



# Proceedings of the Indian Academy of Sciences (Mathematical Sciences)

**Editor**

S G Dani

*Tata Institute of Fundamental Research, Bombay*

**Editorial Board**

S S Abhyankar, *Purdue University, West Lafayette, USA*  
Gopal Prasad, *University of Michigan, Ann Arbor, USA*  
K R Parthasarathy, *Indian Statistical Institute, New Delhi*  
Phoolan Prasad, *Indian Institute of Science, Bangalore*  
M S Raghunathan, *Tata Institute of Fundamental Research, Bombay*  
S Ramanan, *Tata Institute of Fundamental Research, Bombay*  
C S Seshadri, *SPIC Science Foundation, Madras*  
V S Varadarajan, *University of California, Los Angeles, USA*  
S R S Varadhan, *Courant Institute of Mathematical Sciences, New York, USA*  
K S Yajnik, *National Aeronautical Laboratory, Bangalore*

**Editor of Publications of the Academy**

G Srinivasan

*Raman Research Institute, Bangalore*

---

## Subscription Rates (Effective from 1989)

|                                   |        |         |         |
|-----------------------------------|--------|---------|---------|
| <b>All countries except India</b> | 1 year | 3 years | 5 years |
| (Price includes AIR MAIL charges) | US\$75 | \$200   | \$300   |
| <b>India</b>                      | 1 year |         |         |
|                                   | Rs. 75 |         |         |

Annual subscriptions are available for **Individuals** for India and abroad at the concessional rate of Rs. 25/- and \$20/-respectively.

All correspondence regarding subscription should be addressed to **The Circulation Department** of the Academy.

---

### Editorial Office:

Indian Academy of Sciences, C V Raman Avenue,  
P B No. 8005, Bangalore 560 080, India

Telephone: 334 2546  
Telex: 0845-2178 ACAD IN  
Telefax: 91-80-334 6094

© 1994 by the Indian Academy of Sciences. All rights reserved.

The "Notes on the preparation of papers" are printed in the last issue of every volume.



# Proceedings of the Indian Academy of Sciences Mathematical Sciences

Volume 104  
1994

# Proceedings of the Indian Academy of Sciences (Mathematical Sciences)

**Editor**

S G Dani

*Tata Institute of Fundamental Research, Bombay*

**Editorial Board**

S S Abhyankar, *Purdue University, West Lafayette, USA*  
Gopal Prasad, *University of Michigan, Ann Arbor, USA*  
K R Parthasarathy, *Indian Statistical Institute, New Delhi*  
Phoolan Prasad, *Indian Institute of Science, Bangalore*  
M S Raghunathan, *Tata Institute of Fundamental Research, Bombay*  
S Ramanan, *Tata Institute of Fundamental Research, Bombay*  
C S Seshadri, *SPIC Science Foundation, Madras*  
V S Varadarajan, *University of California, Los Angeles, USA*  
S R S Varadhan, *Courant Institute of Mathematical Sciences, New York, USA*  
K S Yajnik, *National Aeronautical Laboratory, Bangalore*

**Editor of Publications of the Academy**

G Srinivasan

*Raman Research Institute, Bangalore*

---

## Subscription Rates (Effective from 1989)

|  |                  |                     |                  |
|--|------------------|---------------------|------------------|
| <b>All countries except India</b><br>(Price includes AIR MAIL charges) | 1 year<br>US\$75 | 3 years<br>\$200    | 5 years<br>\$300 |
| <b>India</b>   | 1 year<br>Rs. 75 | 10 years<br>Rs. 400 |                  |

Annual subscriptions are available for **Individuals** for India and abroad at the concessional rate of Rs. 25/- and \$20/- respectively.

All correspondence regarding subscription should be addressed to **The Circulation Department** of the Academy.

---

### Editorial Office:

Indian Academy of Sciences, C V Raman Avenue,  
P B No. 8005, Bangalore 560 080, India

Telephone: 342546  
Telex: 0845-2178 ACAD IN  
Telefax: 91-812-346094

© 1994 by the Indian Academy of Sciences. All rights reserved.

The "Notes on the preparation of papers" are printed in the last issue of every volume.

# Proceedings of the Indian Academy of Sciences

## Mathematical Sciences

Volume 104, 1994

### CONTENTS

Number 1, February 1994

Special issue dedicated to the memory of Professor K G Ramanathan

Obituary note

On the equation  $x(x + d_1) \dots (x + (k - 1)d_1) = y(y + d_2) \dots (y + (mk - 1)d_2) \dots$   
..... *N Saradha and T N Shorey*

The density of rational points on non-singular hypersurfaces .....  
..... *D R Heath-Brown*

Multiplicative arithmetic of finite quadratic forms over Dedekind rings .....  
..... *Anatoli Andrianov*

Non-surjectivity of the Clifford invariant map .....  
..... *R Parimala and R Sridharan*

Modular forms and differential operators ..... *Don Zagier*

On Fourier coefficients of Maass cusp forms in 3-dimensional hyperbolic  
space ..... *S Raghavan and J Sengupta*

On Zagier's cusp form and the Ramanujan  $\tau$  function .....  
..... *Ashwaq Hashim and M Ram Murty*

Zeta functions of prehomogeneous vector spaces with coefficients related to  
periods of automorphic forms ..... *Fumihiko Sato*

On a problem of G Fejes Toth ..... *R P Bambah and A C Woods*

The number of ideals in a quadratic field ..... *M N Huxley and N Watt*

On the zeros of a class of generalised Dirichlet series - XIV .....  
..... *R Balasubramanian and K Ramachandra*

Local zeta functions of general quadratic polynomials ..... *Jun-ichi Igusa*

Vector bundles as direct images of line bundles .....  
..... *A Hirschowitz and M S Narasimhan*

Finite arithmetic subgroups of  $GL_n$ , III ..... *Yoshiyuki Kitaoka*

|  |  |     |
|--|--|-----|
| Reduction theory over global fields.....   | <i>T A Springer</i>                                    | 207 |
| Symplectic structures on locally compact abelian groups and polarizations<br>..... | <i>R Ranga Rao</i>                                     | 217 |
| Modular equations and Ramanujan's Chapter 16, Entry 29 .....                       | <i>George E Andrews</i>                                | 225 |
| Gaussian quadrature in Ramanujan's Second Notebook .....                           | <i>Richard Askey</i>                                   | 237 |
| Two remarkable doubly exponential series transformations of Ramanujan<br>.....     | <i>Bruce C Berndt and James Lee Hafner</i>             | 245 |
| Kolmogorov's existence theorem for Markov processes in $C^*$ algebras .....        | <i>B V Rajarama Bhat and K R Parthasarathy</i>         | 253 |
| Iterations of random and deterministic functions .....                             | <i>K B Athreya</i>                                     | 263 |
| Existence theory for linearly elastic shells .....                                 | <i>Philippe G Ciarlet</i>                              | 269 |
| Absolutely expedient algorithms for learning Nash equilibria .....                 | <i>V V Phansalkar, P S Sastry and M A L Thathachar</i> | 279 |
| Hierarchic control.....  | <i>J L Lions</i>                                       | 295 |

## Number 2, May 1994

|  |  |     |
|--|--|-----|
| Bertini theorems for ideals linked to a given ideal .....                                | <i>Trivedi Vijaylaxmi</i>                | 305 |
| On the Ramanujan–Petersson conjecture for modular forms of half-integral<br>weight.....  | <i>Winfried Kohnen</i>                   | 333 |
| On composition of some general fractional integral operators.....                        | <i>K C Gupta and R C Soni</i>            | 339 |
| On the absolute matrix summability of Fourier series and some associated<br>series ..... | <i>B K Ray and A K Sahoo</i>             | 351 |
| On absolute summability factors of infinite series .....                                 | <i>Hüseyin Bor</i>                       | 367 |
| Rearrangements of bounded variation sequences .....                                      | <i>Mehmet Ali Sarigöl</i>                | 373 |
| A note on a generalization of Macdonald's identities for $A_1$ and $B_1$ .....           | <i>N Sthanumoorthy and M Tamba</i>       | 377 |
| Combinatorial manifolds with complementarity .....                                       | <i>Basudeb Data</i>                      | 385 |
| Deformations of complex structures on $\Gamma \backslash SL_2(C)$ .....                  | <i>C S Rajan</i>                         | 389 |
| Differential subordinations concerning starlike functions .....                          | <i>S Ponnusamy</i>                       | 397 |
| On the structure of stable random walks.....   | <i>Jon Aaronson</i>                      | 413 |
| $L^1(\mu, X)$ as a complemented subspace of its bidual.....                              | <i>T S S R K Rao</i>                     | 421 |
| Stresses in an elastic plate lying over a base due to strip-loading .....                | <i>Raj Kumar Sharma and Nat Ram Garg</i> | 425 |

**Number 3, August 1994**

|   |   |     |
|---|---|-----|
| The Laplacian on algebraic threefolds with isolated singularities .....                                     | <i>Vishwambhar Pati</i>                               | 435 |
| The Hoffman–Wielandt inequality in infinite dimensions .....  | <i>Rajendra Bhatia and Ludwig Elsner</i>              | 483 |
| Rigidity problem for lattices in solvable Lie groups .....  | <i>A N Starkov</i>                                    | 495 |
| Some remarks on the Jacobian question .....   | <i>Shreeram S Abhyankar</i>                           | 515 |
| On polynomial isotopy of Knot-types .....   | <i>Rama Shukla</i>                                    | 543 |
| Row-reduction and invariants of Diophantine equations .....   | <i>N J Wildberger</i>                                 | 549 |
| Positive values of non-homogeneous indefinite quadratic forms of type (1, 4) .....                          | <i>V C Dumir and Ranjeet Sehmi</i>                    | 557 |
| Extended Kac–Akhiezer formulae and the Fredholm determinant of finite section Hilbert–Schmidt kernels ..... | <i>S Ganapathi Raman and R Vittal Rao</i>             | 581 |
| A proof of Howard’s conjecture in homogeneous parallel shear flows .....                                    | <i>Mihir B Banerjee, R G Shandil and Vinay Kanwar</i> | 593 |

**Number 4, November 1994****Special issue on “Spectral and inverse spectral theory”**

|   |   |     |
|---|---|-----|
| Foreword .....  |   | 597 |
| Scattering theory for Stark Hamiltonians .....                                    | <i>A Jensen</i>                             | 599 |
| $L^p$ -Estimates for Schrödinger operators .....                                  | <i>S Nakamura</i>                           | 653 |
| On $N$ -body Schrödinger operators .....  | <i>H Isozaki</i>                            | 667 |
| A conjecture for some partial differential operators in $L^2(\mathbb{R}^n)$ ..... | <i>Pl. Muthuramalingam</i>                  | 705 |
| The geometry and spectra of hyperbolic manifolds .....                            | <i>P D Hislop</i>                           | 715 |
| Inverse spectral theory for Jacobi matrices and their almost periodicity ....     | <i>A J Antony and M Krishna</i>             | 777 |
| Spectral shift function and trace formula .....                                   | <i>Kalyan B Sinha and<br/>A N Mohapatra</i> | 819 |



# Proceedings of the Indian Academy of Sciences Mathematical Sciences

## Notes on the preparation of papers

Authors wishing to have papers published in the *Proceedings* should send them to

The Editor, Proceedings (Mathematical Sciences), Indian Academy of Sciences,  
C V Raman Avenue, P B No. 8005, Bangalore 560 080, India

OR

Prof. S G Dani, School of Mathematics, Tata Institute of Fundamental Research,  
Homi Bhabha Road, Bombay 400 005, India

Three copies of the paper must be submitted.

The papers must normally present results of original work. Critical reviews of important fields will also be considered for publication. Submission of the typescript will be held to imply that it has not been previously published, it is not under consideration for publication elsewhere and that, if accepted, it will not be published elsewhere. A paper in applied areas will be considered only on the basis of its mathematical content.

### **Typescript:**

Papers should be typed double spaced with ample margin on all sides on white bond paper of size  $280 \times 215$  mm. This also applies to the abstract, tables, figure captions and the list of references which are to be typed on separate sheets.

### **Title page:**

- (1) The title of the paper must be short and contain words useful for indexing.
- (2) The authors' names should be followed by the names and addresses of the institutions of affiliation.
- (3) An *abbreviated running title* of not more than 50 letters and spaces should be given.

### **Abstract:**

Each paper must be accompanied by an abstract describing, in not more than 200 words, the significant results reported in the paper.

### **Keywords:**

Between 3 and 6 keywords must be provided for indexing and information retrieval.

### **The Text:**

The paper should preferably start with an introduction in which the results are placed

## Notes on the preparation of papers

in perspective and some indication of the methods of proof is given.

1. *Markings*: The copy intended for the printer, should be marked appropriately to make it unambiguous. The following conventions may be followed for the purpose of indicating special characters: (A list of the special characters is included at the end)

a) Indicating characters by underlining

|                     |                  |
|---------------------|------------------|
| Italics             | Single underline |
| Bold face           | Wavy underline   |
| Greek               | Blue underline   |
| Fraktur<br>(Gothic) | Red underline    |
| Script              | Green underline  |
| Open face           | Brown underline  |

All parameters in equations are normally printed in italics and numerals in upright typeface.

b) In doubtful cases the position of superscripts and subscripts (exponents and indices) should be clearly marked, for example  $b_j$ ,  $a^{\vee}$ ,  $b^{\infty}$ ,  $b_{x/y}$ , etc.

c) Formulae extending beyond the printed line will be broken by the typesetter at an appropriate place; to avoid any expensive alterations at the proof stage the author may indicate in the manuscript where the formulae should be broken.

d) Various likely ambiguous spots may be explained by pencilled notes in the margin. The following symbols are frequently confused and need to be clarified.

$\circ, o, \cup, O, 0$ ;  $\cup, \bigcup, U$ ;  $\times, x, X, \chi, \kappa$ ;  $\vee, v, \nu$ ;  $\theta, \Theta, \phi, \varphi, \Phi, \emptyset$ ;  $\psi, \Psi$ ;  $\epsilon, \varepsilon$ ;  
 $a, \alpha, \infty$ ;  $B, \beta$ ;  $r, \gamma$ ;  $\sigma, 6$ ;  $+$ ,  $\dagger$ ;  $i, \iota$ ;  $a', a^1$ ; the symbol  $a$  and the indefinite article  $a$ ;  
 also the handwritten Roman letters:  
 $c, C$ ;  $e, \ell$ ;  $I, J$ ;  $k, K$ ;  $o, O$ ;  $p, P$ ;  $s, S$ ;  $u, U$ ;  $v, V$ ;  $w, W$ ;  $x, X$ ;  $z, Z$ .

2. *Notations and style*: Notation and style should be chosen carefully, keeping the printing aspect in mind. The following table indicates preferred forms for various mathematical usages.

| preferred form                        | instead of   | preferred form           | instead of               |
|---------------------------------------|--|--------------------------|--------------------------|
| $A^*, \tilde{b}, \gamma', \nu$ , etc. | $\bar{A}, \hat{b}, \check{\gamma}, \bar{\nu}$ , etc. | $\exp(-(x^2 + y^2)/a^2)$ | $e^{-((x^2 + y^2)/a^2)}$ |
| $\lim \sup$ , $\text{proj } \lim$     | $\lim$ , <u><math>\lim</math></u>                    |                          |                          |
| $f: A \rightarrow B$                  | $A \xrightarrow{\quad} B$                            |                          |                          |
| $\sum_{n=1}^{\infty}$                 | $\sum.$  | $\cos(1/x)$              | $\frac{1}{\cos x}$       |



### **Tables:**

All tables must be numbered consecutively in arabic numerals in the order of appearance in the text. The tables should be self-contained and must have a descriptive title.

### **Figures:**

A figure should be included only when it would be substantially helpful to the reader in understanding the subject matter. The figures should be numbered consecutively in arabic numerals in the order of appearance in the text and the location of each figure should be clearly indicated in margin as "Figure 1 here". The figure captions must be typed on a separate sheet.

Line drawings must be in Indian ink on good quality tracing paper. Lines must be drawn sufficiently thick for reduction to a half or third of the original size (0.3 mm for axes and 0.6 mm for curves are suggested).

### **List of symbols:**

For the convenience of the printers, authors should attach to the manuscript the complete list of symbols identified typographically.

### **References**

References should be cited in the text by serial numbers only (e.g. [3]). They should be listed alphabetically by the author's name (the first author's name in the case of joint authorship) at the end of the paper. In describing each reference the following order should be observed: the author's name followed by initials, title of the article, the name of the journal, the volume number, the year and the page numbers. Standard abbreviations of journal titles should be used. It would be worthwhile to cross-check all references cited in the text with the ones given at the end.

A typical reference to an article in a journal would be like:

- [1] Narasimhan M S and Ramanan S, Moduli of vector bundles on a compact Riemann surface, *Ann. Math.* **89** (1969) 14–51

A References to a book would be on the following lines:

- [8] Royden H, Invariant metrics on Teichmüller space, in: Contributions to analysis (eds) L Ahlfors *et al* (1974) (New York: Academic Press) pp. 393–399

### **Footnotes:**

Footnotes to the text should be avoided if possible. If unavoidable, they should be numbered consecutively, and typed on a separate sheet.

**Proofs:**

The journal is now typeset by computer photocomposition and only a page-proof is supplied to authors. While the process yields better results, it makes corrections at proof stage very difficult and expensive. Deleting a letter or word in the proof can mean resetting the whole line or paragraph. Similarly, addition of a sentence or paragraph might lead to resetting the whole page and sometimes even the whole article.

Authors are requested to prepare the manuscript carefully before submitting it for publication to minimize corrections and alterations in the proof stage which increase publication costs. The proofs sent to the authors together with the reprint order form must be returned to the editorial office *within two days of their receipt*.

**Reprints:**

50 reprints of each article will be supplied free of charge.

Dedicated to the memory of  
**PROFESSOR K G RAMANATHAN**



Professor K G Ramanathan

**Professor K G Ramanathan**  
**(1920–1992)**

He was small of build but had a big influence on the post-independence Indian mathematical scene. Despite the legacy of the legendary Srinivasa Ramanujan and several other mathematicians of high standing early in this century, pursuit of mathematics had remained rather weak in India till the fifties. Professor K G Ramanathan was one of the few people responsible for the fortification which has put India firmly back on the international mathematical map. He not only was himself a front-ranking mathematician of international reputation, but also contributed a great deal to the emergence of a strong mathematical base at the Tata Institute of Fundamental Research as also to the overall development of research and teaching of mathematics in India and, to an extent, even beyond our shores.

Kollagunta Gopalaiyer Ramanathan was born on 13 November 1920 in Hyderabad in South India. He got his B. A. from Osmania University (1940), M. A. from the University of Madras (1942) and Ph.D. from Princeton University in 1951. At Princeton he came under the influence of the great mathematician Carl L Siegel, whose deep imprint is noticable on his later career. Soon after his doctorate he joined the Tata Institute of Fundamental Research where he teamed up with Professor K Chandrasekharan in building the School of Mathematics of the Institute and in particular the Number Theory group. His abiding enthusiasm, professional expertise, meticulous style of teaching, tireless working ability and good taste in mathematics played a pivotal role in the development of the School. Not content with his role in building an excellent centre in pure mathematics, he embarked on setting up a centre for application of mathematics, in Bangalore, in collaboration with the Indian Institute of Science. With the help of his contacts with many eminent applied mathematicians around the world, he nurtured the IISc–TIFR programme which has over the years given rise to an active and internationally successful group in applied mathematics.

As a mathematician Professor Ramanathan was well recognized for his achievements in Number Theory, especially the analytic and arithmetic theory of quadratic forms over division algebras with involution. Applying his results on quadratic forms he constructed, in an important paper, infinitely many classes of mutually incommensurable discrete subgroups of the first kind in classical semisimple groups and, following that, in another remarkable paper, settled the question of maximality of discrete subgroups of arithmetically defined classical groups, generalising certain results of Hecke and Maass. A conjecture due to A Oppenheim on values of indefinite real quadratic forms at integral points, which was settled only recently by G A Margulis, was another of his favourite problems and he made a significant contribution, jointly with S Raghavan, pertaining to it. During his last years he had engaged himself actively in studying and expounding the so-called Lost Notebooks of Ramanujan and obtained fruitful extensions of Ramanujan's work on singular values of certain modular functions, Rogers–Ramanujan continued fractions and hypergeometric series.

Professor Ramanathan was the Editor of the *Journal of the Indian Mathematical Society* from 1969 to 1982. The high standard of publication maintained by the Journal during the years, despite the all too familiar difficulties in this respect especially in India, is a testimony to Professor Ramanathan's commitment to quality. He was a member of the Editorial Board of *Acta Arithmetica* for nearly three decades. He was a colourful personality with ready wit and clear thoughts, which he expressed unhesitatingly. Several universities and other academic institutions benefited from his characteristically frank advice on various matters.

As was to be expected, he received numerous honours. He was a Fellow of the Indian Academy of Sciences and the Indian National Science Academy and Founder Fellow of the Maharashtra Academy of Sciences. He served as the President of the Indian Mathematical Society and Life President of the Bombay Mathematical Colloquium. He was awarded the Shanti Swarup Bhatnagar Prize, the Jawaharlal Nehru Fellowship, the Homi Bhabha Medal (of INSA) and the Padma Bhushan.

His early death on 10 May 1992, following poor health for several years during which he continued to be actively involved with mathematics, has taken away from our midst a leader championing the cause of good mathematics in a multitude of ways. It is a grievous loss not only to his family, his students and others whose lives were shaped by him in varying degrees, but to the mathematical community as a whole.

**On the equation  $x(x + d_1) \dots (x + (k - 1)d_1) = y(y + d_2) \dots (y + (mk - 1)d_2)$**

N SARADHA and T N SHOREY

School of Mathematics, Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400 005, India

Dedicated to the memory of Professor K G Ramanathan

**Abstract.** For given positive integers  $m \geq 2, d_1$  and  $d_2$ , we consider the equation of the title in positive integers  $x, y$  and  $k \geq 2$ . We show that the equation implies that  $k$  is bounded. For a fixed  $k$ , we give conditions under which the equation implies that  $\max(x, y)$  is bounded.

**Keywords.** Exponential diophantine equations; arithmetic-geometric mean.

## 1. Introduction

For positive integers  $m \geq 2, d_1$  and  $d_2$ , we consider the equation

$$x(x + d_1) \dots (x + (k - 1)d_1) = y(y + d_2) \dots (y + (mk - 1)d_2) \quad (1)$$

in integers  $x > 0, y > 0$  and  $k \geq 2$ . Equation (1) with  $d_1 = d_2$  was considered in [4] and [5]. It was shown in [5, Corollary 2] that equation (1) with  $d_1 = d_2 = d$  and  $m \geq 2$  implies that  $\max(x, y, k)$  is bounded by an effectively computable number depending only on  $m$  and  $d$ . In this paper, we extend this result as follows:

**Theorem 1.** *There exists an effectively computable number  $C$  depending only on  $d_1$  and  $d_2$  such that equation (1) with  $m = 2$  implies that either*

$$\max(x, y, k) \leq C$$

or

$$k = 2, d_1 = 2d_2^2, x = y^2 + 3d_2y.$$

On the other hand, we observe that equation (1) with  $m = 2$  is satisfied whenever the latter possibility holds.

**Theorem 2.** *Let  $m > 2$ . Assume that equation (1) is satisfied. Then*

- (a)  *$k$  is bounded by an effectively computable number  $C_1$  depending only on  $m, d_1$  and  $d_2$ .*
- (b) *Let  $k \leq C_1$ . There exists an effectively computable number  $C_2$  depending only on  $m, d_1$  and  $d_2$  such that either*

$$\max(x, y) \leq C_2 \quad (2)$$

or

*$d_1/d_2^m$  is a product of  $m$  distinct positive integers composed of primes not exceeding  $m$ .*

(c) Let  $k \leq C_1$ . Then, either (2) holds or

$$m \geq \alpha(k) \quad (3)$$

where

$$\alpha(k) = \begin{cases} 14 & \text{for } 2 \leq k \leq 7 \\ 50 & \text{for } k = 8 \\ \exp\{k \log k - (1.25475)k - \log k + 1.56577\} & \text{for } k \geq 9. \end{cases} \quad (4)$$

We observe from (3) and (4) that  $m \geq 14$  for  $k \geq 2$  and  $m \geq 2568$  for  $k \geq 9$ ,  $m \geq 17010$  for  $k \geq 10$ ,  $m \geq 125804$  for  $k \geq 11$ . Thus, we observe from Theorem 2(a) and Theorem 2(c) that equation (1) with  $3 \leq m \leq 13$  implies that  $\max(x, y, k)$  is bounded by an effectively computable number depending only on  $d_1$  and  $d_2$ . This is also the case whenever equation (1) with  $m < 2568$  and  $k \geq 9$  is valid. More generally, equation (1) with  $m > 2$  and

$$k \geq \max(10, (21 \log m)/20)$$

implies that  $\max(x, y, k)$  is bounded by an effectively computable number depending only on  $m, d_1$  and  $d_2$ . Finally, we remark that Theorem 2(b) is applied in the proof of Theorem 2(c).

## 2. Lemmas

In this section, we prove lemmas for the proofs of the theorems. The lemmas are more general than required and we hope that they may be of independent interest. We start with the following extension of [5, Lemma 1]. We write  $N$  for a positive number given by

$$N^2 = (m-1)k \text{ with } m \geq 2. \quad (5)$$

*Lemma 1.* Let  $\varepsilon > 0$  and  $m \geq 2$ . There exists an effectively computable number  $C_3$  depending only on  $\varepsilon$  such that equation (1) with  $k \geq C_3$  and

$$x \geq d_1 \quad (6)$$

implies that

$$\log x \geq \left(\frac{1}{2} - \varepsilon\right)N. \quad (7)$$

*Proof.* We may assume that  $k$  exceeds a sufficiently large effectively computable number depending only on  $\varepsilon$ . Then, by equation (1) and (6), we have

$$(mk-1)! d_2^{mk-1} \leq (k!)x^k$$

which implies that

$$x \geq e^{-1} k^{m-1} d_2^{(mk-1)/k}. \quad (8)$$

Thus

$$x \geq (d_1 d_2)^{1/2}, \quad (9)$$



i.e.  $x < d_2$  which contradicts (8). If all primes not exceeding  $N$  divide  $d_1 d_2$ , we observe from (9) and Prime Number Theory that

$$\log x \geq \frac{1}{2} \log(d_1 d_2) \geq (1 - \varepsilon)N/2. \quad (10)$$

On the other hand, if there exists a prime  $p \leq N$  such that  $p \nmid d_1 d_2$ , then we argue  $p$ -adically as in [5, Lemma 1] to obtain

$$\frac{k}{p}(m-1) < \frac{\log(x + (k-1)d_1)}{\log p} + 2. \quad (11)$$

Now, we combine (11) and (5) for deriving that

$$\log(x + (k-1)d_1) \geq (1 - \varepsilon)N/2$$

which, together with (5) and (6), implies (7).  $\square$

As an immediate consequence of Lemma 1, we obtain the following extension of [5, Corollary 3].

### COROLLARY 1

Let  $\varepsilon > 0$  and  $m \geq 2$ . If (1) and (6) hold, then

$$\log\left(y + \left(\frac{mk-1}{2}\right)d_2\right) \geq \left(\frac{1}{2} - \varepsilon\right)N/m \text{ for } k \geq C_3. \quad (12)$$

*Proof.* We apply arithmetic-geometric mean to the right hand side of (1) to derive (12) from (7) as in the proof of [5, Corollary 3].  $\square$

Let  $B_j = B_j(m, k)$  be given by [4, (3)–(5)]. We prove

*Lemma 2.* Let  $\varepsilon > 0$  and  $m \geq 2$ . The equation (1) with

$$d_1 k^{m+1} \leq x^{1/2} \quad (13)$$

and

$$d_2 \leq y^{(1-\varepsilon)/(m+1)} \quad (14)$$

implies that either

$$x_1 = y_2^m + B_1 d_2 y_2^{m-1} + \dots + B_m d_2^m - \left(\frac{k+1}{2}\right)d_1 \quad (15)$$

where

$$x_1 = x - d_1, \quad y_2 = y - d_2 \quad (16)$$

or

$$\max(x, y, k) \leq C_4 \quad (17)$$

for some effectively computable number  $C_4$  depending only on  $\varepsilon$  and  $m$ .

*Proof.* Let  $0 < \varepsilon < 1$  and  $m \geq 2$ . We assume (1) with (13) and (14). Then, we observe that  $d_1 < x$ ,  $d_2 < y$  and  $x_1, y_2$  are positive integers. By (1) and (16), we have

$$(x_1 + d_1) \dots (x_1 + kd_1) = (y_2 + d_2) \dots (y_2 + mkd_2). \quad (18)$$

We denote by  $c_1, c_2, c_3$  and  $c_4$  effectively computable positive numbers depending only on  $\varepsilon$  and  $m$ . We may assume that  $y_2 \geq c_1$  with  $c_1$  sufficiently large, otherwise we derive from (12), (5), (14) and (1) that  $\max(x, y, k) \leq c_2$ . Next, we observe from Corollary 1 that

$$\log(y_2 + (mk - 1)d_2) \geq c_3 k^{1/2}. \quad (19)$$

Also, we observe from Lemma 1 that

$$\log x_1 \geq c_4 k^{1/2}. \quad (20)$$

Now, we follow the proof of [5, §3]. We define  $A_j(m, k), B_j = B_j(m, k)$  and  $H_j(m, k)$  as in [4, (2)–(5)]. Further, we define

$$F_{d_1}(x_1, k) = (x_1 + d_1) \dots (x_1 + kd_1),$$

$$F_{d_2}(y_2, m, k) = (y_2 + d_2) \dots (y_2 + mkd_2)$$

and

$$\Lambda_{d_2} = \Lambda_{d_2}(y_2, m, k) = y_2^m + B_1 d_2 y_2^{m-1} + \dots + B_m d_2^m. \quad (21)$$

When  $d_1 = d_2 = d$ , these definitions coincide with the corresponding definitions in [5].

By applying arithmetic-geometric mean to the left hand side of (1), we obtain

$$F_{d_1}(x_1, k) < \left( x_1 + \frac{k+1}{2} d_1 \right)^k.$$

Now, we use (18), (19), (20) and we argue as in the proof of [4, Lemma 5] to obtain

$$F_{d_2}(y_2, m, k) < (\Lambda_{d_2} + (4k^{2m-1})^{-1})^k,$$

$$F_{d_2}(y_2, m, k) > (\Lambda_{d_2} - (2k^{2m-1})^{-1})^k$$

and

$$F_{d_1}(x_1, k) > \left( x_1 + \frac{k+1}{2} d_1 - (4k^{2m-1})^{-1} \right)^k.$$

Finally, we utilise these estimates and [4, Lemma 3] to conclude that equation (1) implies that

$$x_1 = \Lambda_{d_2} + fd_1, f = -(k+1)/2, \quad (22)$$

which, by (21), coincides with (15).  $\square$

*Lemma 3.* Let  $\varepsilon > 0$  and  $m > 2$ . There exist effectively computable numbers  $C_5, C_6$  and  $C_7$  depending only on  $\varepsilon$  and  $m$  such that equation (1) with  $\max(x, y, k) \geq C_5$ , (13) and (14) implies that  $m \geq 14$ ,  $k \leq C_6$  and

$$\mu d_2^m = v d_1 \quad (23)$$

$$\max(\mu, \nu) \leq C_7. \quad (24)$$

*Proof.* We may assume that  $C_5 > C_4$  so that we derive from Lemma 2 that (15) is valid. Further, we re-write (15) as (22) and we substitute (22) in the left hand side of equation (18) to obtain

$$F_{d_1}(x_1, k) = \Lambda_{d_2}^k + a_2(f, k)d_1^2\Lambda_{d_2}^{k-2} + \dots + a_k(f, k)d_1^k \quad (25)$$

where  $a_i(f, k)$  with  $1 \leq i \leq k$  are given by [4, (44) and (45)]. Now, we substitute (21) in (25) for writing

$$F_{d_1}(x_1, k) = \sum_{j=0}^{mk} T_{j, d_1, d_2}(m, k) d_2^j y_2^{mk-j}$$

where

$$T_{j, d_1, d_2}(m, k) = \begin{cases} H_j(m, k) \text{ for } 0 \leq j < 2m, \\ H_j(m, k) + a_2(f, k)d_1^2 d_2^{-2m} H_{j-2m}(m, k-2) + \dots \\ \quad + a_h(f, k)d_1^h d_2^{-hm} H_{j-hm}(m, k-h) \text{ for} \\ \quad hm \leq j < (h+1)m \text{ and } 2 \leq h < k, \\ B_m^k + a_2(f, k)d_1^2 d_2^{-2m} B_m^{k-2} + \dots + \\ \quad a_k(f, k)d_1^k d_2^{-km} \text{ for } j = mk \end{cases}$$

Proceeding as in the proof of [4, (57) and (58)], we derive that

$$H_j(m, k) = A_j(m, k) \text{ for } 0 \leq j \leq 2m \quad (26)$$

and

$$(H_{2m}(m, k) - A_{2m}(m, k))d_2^{2m} = \frac{k(k-1)(k+1)}{24} d_1^2. \quad (27)$$

From the explicit calculations using the method described by Glesser in [3, Appendix], we derive that

$$H_j(m, k) - A_j(m, k) > 0 \text{ for } k \geq 2, m \leq 13 \quad (28)$$

where  $j = m + 1$  if  $m$  is odd and  $j = m + 2$  if  $m$  is even. By (26) and (28), we derive that  $m \geq 14$ .

Since  $m > 2$ , we apply a result of Balasubramanian [4, Appendix] to obtain from (26) that  $k$  is bounded by an effectively computable number depending only on  $\varepsilon$  and  $m$ . Finally, we take square roots on both the sides of (27) to obtain (23) satisfying (24).  $\square$

If  $m > 2$ , we show that the hypothesis (13) is not required whenever equation (1) with  $d_1 = d_2$  is satisfied. If (13) is not valid, we observe from [5, (7)] that

$$x^{1/10} < k^{m+1}$$

which, by [5, Lemma 1], implies that  $\max(x, y, k)$  is bounded by an effectively

$$\mu^k x'(x' + 1) \dots (x' + (k - 1)) = v^k y'(y' + 1) \dots (y' + (mk - 1)) \quad (29)$$

where  $x' = x/d_1$ , and  $y' = y/d_2$ . Next, in view of (24) and  $k \leq C_6$ , we apply the theorem of Faltings (under suitable assumptions) to equation (29) for concluding that there are only finitely many possibilities for  $x, y$  satisfying (1). For deriving this assertion from equation (1) with (23), (24) and  $k \leq C_6$ , we shall not utilise the theorem of Faltings as it is non-effective. We shall follow an elementary approach which is valid under certain restrictions.

Let  $g = \gcd(d_1, d_2^m)$  and  $f(X)$  be a positive real valued function of a positive real variable  $X$  satisfying

$$\lim_{X \rightarrow \infty} f(X) = \infty.$$

We derive from Lemma 3 the following result.

*Lemma 4. Let  $m > 2$  and  $\theta > 0$ . The equation (1) with (13), (14) and*

$$g \leq \theta \max\left(\frac{d_1}{f(d_1)}, \frac{d_2^m}{f(d_2)}\right) \quad (30)$$

*implies that*

$$\max(d_1, d_2, k) \leq C_8 \quad (31)$$

*where  $C_8$  is an effectively computable number depending only on  $\varepsilon, m, f$  and  $\theta$ .*

*Proof.* We write  $C_9, C_{10}$  and  $C_{11}$  for effectively computable numbers depending only on  $\varepsilon, m, f$  and  $\theta$ . By Lemma 3, we conclude that

$$k \leq C_9 \quad (32)$$

and (23) with (24) is valid. We divide both the sides of (23) by  $g$  to derive from (24) that

$$\max\left(\frac{d_1}{g}, \frac{d_2^m}{g}\right) \leq C_7. \quad (33)$$

By (33) and (30), we observe that

$$\min(f(d_1), f(d_2)) \leq \theta C_7.$$

Now, by the definition of  $f$ , we obtain

$$\min(d_1, d_2) \leq C_{10}$$

which, together with (27) and (32), implies that  $\max(d_1, d_2) \leq C_{11}$ . □

The assumption (30) is satisfied whenever one of the following conditions holds. (The choice of  $\theta$  and  $f$  is given in the brackets)

$$(i) \quad d_1 \text{ fixed} \quad (\theta = f(d_1))$$

- (ii)  $d_2$  fixed  $(\theta = f(d_2))$
- (iii)  $\gcd(d_1, d_2) = 1$   $(\theta = 1, f(X) = X)$
- (iv)  $d_1 = d_2$   $(\theta = 1, f(X) = X)$
- (v)  $d_1 \leq d_2^m / \log(d_2 + 1)$   $(\theta = 1, f(X) = \log(X + 1))$
- (vi)  $d_2^m \leq d_1 / \log(d_1 + 1)$   $(\theta = 1, f(X) = \log(X + 1))$

Therefore, equation (1) with  $m > 2$ , (13) and (14) implies (31) if at least one of the assumption (i)–(vi) holds. As remarked earlier, the assumption (13) is not required whenever  $m > 2$  and (iv) is valid. In the next section, we prove Theorem 2(a) by showing that the assumptions (13) and (14) are not needed whenever  $d_1$  and  $d_2$  are fixed.

### 3. Proof of Theorem 2(a)

We may suppose that  $y$  exceeds a sufficiently large effectively computable number depending only on  $m, d_1$  and  $d_2$ , otherwise the assertion of Theorem 2(a) follows immediately from (12) and (5). Then (14) is satisfied and (13) is a consequence of (7). Now, as remarked at the end of the previous section, we conclude the assertion of Theorem 2(a).  $\square$

### 4. Proofs of Theorem 2(b) and Theorem 2(c)

In this section, we shall always assume that equation (1) with

$$m > 2, \quad k \leq C_1 \tag{34}$$

is satisfied. Then, by equation (1), we may assume that  $y_2 > y'$  where  $y'$  is a sufficiently large effectively computable number depending only on  $k, m, d_1, d_2$  and  $y_2$  is given by (16), otherwise Theorem 2(b) and Theorem 2(c) follow immediately from (34). Then  $x_1$ , given by (16), is positive and (18) is valid. Also, we observe that (13) and (14) are satisfied. We put

$$D = d_1 / d_2^m, \tag{35}$$

$$\phi(Y) = Y^m + B_1 d_2 Y^{m-1} + \dots + B_m d_2^m - \left( \frac{k+1}{2} \right) d_1, \tag{36}$$

$$L(X, Y) = (X + d_1) \dots (X + k d_1) - (Y + d_2) \dots (Y + m k d_2) \tag{37}$$

and

$$l(Y) = L(\phi(Y), Y). \tag{38}$$

Now, we apply Lemma 2 and (18) to suppose that  $l(y_2) = 0$ . Then, since  $y'$  is sufficiently large, we derive from (34) that

$$l(Y) \equiv 0. \tag{39}$$

By (36), (37), (38) and (39), we obtain pairwise distinct integers

$$1 \leq \lambda_{i,j} \leq m k, \quad 1 \leq i \leq k, \quad 1 \leq j \leq m, \tag{40}$$

such that

$$\phi(Y) + id_1 \equiv (Y + \lambda_{i,1}d_2) \dots (Y + \lambda_{i,m}d_2) \text{ for } 1 \leq i \leq k. \quad (41)$$

We observe that (40) covers all the integers in the interval  $[1, mk]$ . There is no loss of generality in assuming that each  $m$ -tuple  $\{\lambda_{i,1}, \dots, \lambda_{i,m}\}$  is such that

$$\lambda_{i,1} < \dots < \lambda_{i,m} \text{ for } 1 \leq i \leq k. \quad (42)$$

Let  $\{\lambda_{i_0,1}, \dots, \lambda_{i_0,m}\}$  be the  $m$ -tuple containing 1. Then, we observe from (42) that  $\lambda_{i_0,1} = 1$ . Further, we derive from (41) that

$$(i - i_0)d_1 \equiv (Y + \lambda_{i,1}d_2) \dots (Y + \lambda_{i,m}d_2) - (Y + \lambda_{i_0,1}d_2) \dots (Y + \lambda_{i_0,m}d_2) \text{ for } 1 \leq i \leq k, i \neq i_0. \quad (43)$$

By putting  $Y = -\lambda_{i_0,1}d_2 = -d_2$  in (43), we get

$$(i - i_0)d_1 = (\lambda_{i,1} - 1) \dots (\lambda_{i,m} - 1)d_2^m \text{ for } 1 \leq i \leq k, i \neq i_0. \quad (44)$$

We observe from (44) that

$$i_0 = 1. \quad (45)$$

It follows from  $B_1 = m(mk + 1)/2$ , (36) and (41) that

$$m(mk + 1)/2 = \sum_{j=1}^m \lambda_{i,j} \text{ for } 1 \leq i \leq k. \quad (46)$$

Further, we set

$$\Delta_i = (\lambda_{i,1} - 1) \dots (\lambda_{i,m} - 1) \text{ for } 2 \leq i \leq k, \quad (47)$$

$$\Delta_1 = (\lambda_{1,2} - 1) \dots (\lambda_{1,m} - 1) \quad (48)$$

and

$$\Omega = \prod_{i=2}^k \Delta_i. \quad (49)$$

Now, we observe from (47), (48), (49), (40), (44) and (45) that

$$\Delta_1 \Omega = (mk - 1)! \quad (50)$$

and

$$\Omega = (k - 1)! D^{k-1}. \quad (51)$$

*Proof of Theorem 2(b).* By (44) with  $i = 2$ , (45) and (35), we conclude that  $D$  is a product of  $m$  distinct positive integers. Therefore, it suffices to show that every prime divisor of  $D$  is at most  $m$ . We set

$$\psi(Z) = Z^m + B_1 Z^{m-1} + \dots + B_m - \left( \frac{k+1}{2} \right) D \quad (52)$$

By (36) and (52),

$$\phi(Y) \equiv \psi(Y) \pmod{D}$$

Further, by (41) and (53), it follows that

$$\psi(Z) + iD \equiv (Z + \lambda_{i,1}) \dots (Z + \lambda_{i,m}) \text{ for } 1 \leq i \leq k. \quad (54)$$

Since  $D$  is an integer, we observe from (54) that  $\psi(Z)$  is a polynomial of degree  $m$  with integer coefficients.

Let  $p$  be a prime divisor of  $D$ . By (44) with  $i = 2$  and (45), we observe that  $p < mk$ . Then, we derive from (54) that

$$\psi(-v) \equiv 0 \pmod{p} \text{ for } 1 \leq v \leq p. \quad (55)$$

This implies that  $p \leq m$ , since  $\psi(Z) \equiv 0 \pmod{p}$  has at most  $m$  incongruent solutions mod  $p$ .  $\square$

For an integer  $v > 1$ , we write  $P(v)$  for the greatest prime factor of  $v$  and we put  $P(1) = 1$ . The letter  $p$  denotes always a prime number. For the proof of Theorem 2(c), we require the following results from Prime Number Theory. The first result is a sharpening, due to Hanson [1], of a theorem of Sylvester.

*Lemma 5. For positive integers  $k \geq 2$  and  $n > k$ , either*

$$P(n(n+1) \dots (n+k-1)) > 3k/2$$

*or*

$$(n, k) \in \{(3, 2), (8, 2), (6, 5)\}.$$

The second result is due to Rosser and Schoenfeld [2, p 65–70.] on estimates for some well-known functions in Prime Number Theory. Let

$$\pi(x) = \sum_{p \leq x} 1$$

$$\vartheta(x) = \sum_{p \leq x} \log p$$

and

$$E = -\gamma - \sum_{n=2}^{\infty} \sum_p (\log p)/p^n$$

where  $\gamma$  is Euler's constant. Then

*Lemma 6. For  $x \geq 2$ , we have*

$$\pi(x) > x/(\log x) \text{ for } x \geq 17, \quad (56)$$

$$\pi(x) < 13x/(10 \log x), \quad (57)$$

$$\sum_{p \leq x} (\log p)/p > \log x + E - 1/(2 \log x), \quad (58)$$

$$\sum_{p \leq x} (\log p)/p < \log x + E + 1/(\log x) \text{ for } x \geq 32, \quad (59)$$

$$\vartheta(x) > x(1 - 1/(\log x)) \text{ for } x \geq 41, \quad (60)$$

$$\vartheta(x) < x(1 + 1/(2 \log x)). \quad (61)$$

By taking  $y'$  sufficiently large, we derive from Lemma 3 that

$$m \geq 14. \quad (62)$$

Further, we apply Lemma 5 to sharpen (62) as follows.

*Lemma 7. We have*

$$m > k. \quad (63)$$

*Proof.* By (62), we may assume that

$$k \geq 13. \quad (64)$$

We denote by  $\mu_1 < \mu_2 < \dots < \mu_s$  the elements of  $\{\lambda_{1,2} - 1, \dots, \lambda_{1,m} - 1\}$  which are greater than  $k$ . We observe that

$$0 \leq s \leq m - 1. \quad (65)$$

By writing  $\mu_0 = k$  and  $\mu_{s+1} = mk$ , we divide

$$(k, mk) - \{\mu_1, \dots, \mu_s\}$$

into  $(s + 1)$  disjoint intervals

$$(\mu_j, \mu_{j+1}) \text{ for } 0 \leq j \leq s.$$

Then, we find  $J$  with  $0 \leq J \leq s$  satisfying

$$\mu_{J+1} - \mu_J - 1 \geq (mk - k - s - 1)/(s + 1). \quad (66)$$

By (66), (65), (62) and (64), we derive that

$$\mu_{J+1} - \mu_J - 1 \geq (13k/14) - 1 > 2k/3. \quad (67)$$

Now, we derive from (67) and Lemma 5 that the interval  $(\mu_J, \mu_{J+1})$  contains an integer  $\mu$  divisible by a prime  $> k$ . Further, we observe from (49) that  $\mu$  divides  $\Omega$ . Therefore, we conclude from (51) and Theorem 2(b) that

$$k < P(\mu) \leq P(D) \leq m. \quad \square$$

*Lemma 8. For  $k \geq 8$ , we have*

$$\log m > k - \log k - 2. \quad (68)$$

*Proof.* By (51), (63) and Theorem 2(b), we derive that

$$w(\Omega) \leq \pi(m) \quad (69)$$

where  $w(\Omega)$  denotes the number of distinct prime divisors of  $\Omega$ . On the other hand, we observe from (50) and (48) that

$$w(\Omega) > \pi(mk) - m \quad (70)$$



Now, we apply (56) and (57) in (71) for deriving that

$$\log m > k - \log k - \left( \frac{13}{10} + \frac{13 \log k}{10 \log m} \right). \quad (72)$$

By (72) and (63), we have

$$\log m > k - \log k - 2.6. \quad (73)$$

Then, since  $k \geq 8$ , we observe from (73) that  $m \geq 28$ . Now, we derive from (72) that

$$\log m > k - \log k - 2.15.$$

Repeating this process two more times, we obtain (68).  $\square$

*Proof of Theorem 2(c).* By (62) and (68), we may assume that  $k \geq 9$ . Then, we observe from (68) that  $m \geq 115$ . The proof depends on comparing an upper and lower bound for  $\Delta_1$ . By (48), we obtain

$$\Delta_1 < \lambda_{1,2} \dots \lambda_{1,m}$$

which, by arithmetic – geometric mean and (46), implies that

$$\Delta_1 < \left( \frac{m(mk+1)}{2(m-1)} \right)^{m-1} < e \left( \frac{mk+1}{2} \right)^{m-1}. \quad (74)$$

By (50), (51), (63) and a consequence  $P(D) \leq m$  of Theorem 2(b), we conclude that

$$\log \Delta_1 \geq \sum_{m < p \leq mk} \text{ord}_p((mk-1)!) \log p.$$

Therefore

$$\log \Delta_1 \geq (mk-1) \sum_{m < p \leq mk} \frac{\log p}{p} - \mathfrak{O}(mk-1) + \mathfrak{O}(m). \quad (75)$$

Now, we apply (58), (59), (61) and (60) in (75) for deriving

$$\log \Delta_1 > (mk-1)(\log k - 2/(\log m)) - mk + m + 1 - m/(\log m). \quad (76)$$

Next, we combine (76) and (74) to obtain

$$\log m > k \log k - k - \log k - \frac{2k+1}{\log m} + \log 2 + 1. \quad (77)$$

Then, we observe from (77) and  $m \geq 115$  that

$$\log m > k \log k - (1.4216)k - \log k + 1.48239.$$

Repeated applications of (77), as in the proof of Lemma 8, yield

$$\log m > k \log k - (1.25475)k - \log k + 1.56577. \quad \square$$

## 5. Proof of Theorem 1.

Let  $m = 2$ . Suppose that equation (1) is satisfied. As earlier, we may assume that  $y$  exceeds a sufficiently large effectively computable number depending only on  $d_1$  and  $d_2$ . Further, the inequalities (13) and (14) are valid. Consequently, we conclude (15). Next, we argue as in the proof of Lemma 3, for deriving (27). We calculate

$$H_4(2, k) - A_4(2, k) = (4k^5 - 5k^3 + k)/90. \quad (78)$$

By (27) and (78), we find that

$$D^2 = (d_1/d_2^2)^2 = 4(4k^2 - 1)/15. \quad (79)$$

In particular, we observe that  $k$  is bounded by an effectively computable number depending only on  $d_1$  and  $d_2$ . Further, as in the proof of Theorem 2(b), we show that  $D$  is an integer satisfying  $P(D) = 2$ . Then, we conclude from (79) that  $D = k = 2$ , which together with (15), implies that  $x = y^2 + 3d_2y$ .  $\square$

## Acknowledgements

The authors are thankful to Professor K Ramachandra for drawing their attention to equation (1).

## References

- [1] Hanson D, On a theorem of Sylvester and Schur, *Can. Math. Bull.* **16** (1973) 195–199
- [2] Rosser B and Schoenfeld L, Approximate formulas for some functions of prime numbers, *Illinois. J. Math.* **6** (1962) 64–94
- [3] Saradha N and Shorey T N, The equations  $(x + 1) \dots (x + k) = (y + 1) \dots (y + mk)$  with  $m = 3, 4$ , *Indagationes Math.* **2(4)** (1991) 489–510
- [4] Saradha N and Shorey T N, On the equation  $(x + 1) \dots (x + k) = (y + 1) \dots (y + mk)$ , *Indagationes Math.* **3** (1992) 79–90
- [5] Saradha N and Shorey T N, On the equation  $x(x + d) \dots (x + (k - 1)d) = y(y + d) \dots (y + (mk - 1)d)$ , *Indagationes Math.* **3** (1992) 237–242

# The density of rational points on non-singular hypersurfaces

D R HEATH-BROWN

Magdalen College, Oxford OX14AU, England

Dedicated to the memory of Professor K G Ramanathan

**Abstract.** Let  $F(\mathbf{x}) = F[x_1, \dots, x_n] \in \mathbb{Z}[x_1, \dots, x_n]$  be a non-singular form of degree  $d \geq 2$ , and let

$$N(F, X) = \#\{\mathbf{x} \in \mathbb{Z}^n; F(\mathbf{x}) = 0, |\mathbf{x}| \leq X\},$$

where

$$|\mathbf{x}| = \max_{1 \leq r \leq n} |x_r|.$$

It was shown by Fujiwara [4] [Upper bounds for the number of lattice points on hypersurfaces, *Number theory and combinatorics, Japan, 1984*, (World Scientific Publishing Co., Singapore, 1985)] that  $N(F, X) \ll X^{n-2+2/n}$  for any fixed form  $F$ . It is shown here that the exponent may be reduced to  $n-2+2/(n+1)$ , for  $n \geq 4$ , and to  $n-3+15/(n+5)$  for  $n \geq 8$  and  $d \geq 3$ . It is conjectured that the exponent  $n-2+\varepsilon$  is admissible as soon as  $n \geq 3$ . Thus the conjecture is established for  $n \geq 10$ . The proof uses Deligne's bounds for exponential sums and for the number of points on hypersurfaces over finite fields. However a composite modulus is used so that one can apply the ' $q$ -analogue' of van der Corput's AB process.

**Keywords.** Rational points; hypersurface; counting function; multiple exponential sum; Deligne's bounds; singular locus.

## 1. Introduction

Let  $F(\mathbf{x}) = F[x_1, \dots, x_n] \in \mathbb{Z}[x_1, \dots, x_n]$  be a non-zero form of degree  $d$ . We shall be concerned here with bounds for the number

$$N(F, X) = \#\{\mathbf{x} \in \mathbb{Z}^n; F(\mathbf{x}) = 0, |\mathbf{x}| \leq X\},$$

where

$$|\mathbf{x}| = \max_{1 \leq r \leq n} |x_r|.$$

It is trivial that

$$N(F, X) \ll_F X^{n-1}.$$

Moreover it is clear that  $N(F, X) \gg_F X^{n-1}$  whenever  $F$  has a rational linear factor. However in all other cases one has

$$N(F, X) \ll_F X^{n-3/2} \log X. \quad (1)$$

This follows from Lemma 15 of Heath-Brown [7], where the result is deduced from estimates of Cohen [2]. It is reasonable to conjecture rather more:

**CONJECTURE.** Let  $F(\mathbf{x}) = F[x_1, \dots, x_n] \in \mathbb{Z}[x_1, \dots, x_n]$  be a non-zero form with no rational factor. Then

$$N(F, X) \ll_{F, \varepsilon} X^{n-2+\varepsilon} \quad (2)$$

for any  $\varepsilon > 0$ .

When  $F$  is non-singular one has the bound  $N(F, X) \ll_F X^{n-2+2/n}$  of Fujiwara [4], which approximates to (2) as  $n$  tends to infinity. Our first result gives a slight sharpening of Fujiwara's result and is of significance for small values of  $n$ .

**Theorem 1.** Let  $F(\mathbf{x}) = F[x_1, \dots, x_n] \in \mathbb{Z}[x_1, \dots, x_n]$  be a non-singular form of degree  $d \geq 2$ , and suppose that  $n \geq 4$ . Then

$$N(F, X) \ll_F X^{n-2+2/(n+1)}.$$

In fact the proof shows that

$$N(F, X) \ll_F X^{3/2} \log X$$

for  $n = 3$ . However this estimate, without the factor  $\log X$ , can be obtained by a more elementary route. Indeed, using Falting's Theorem, one can show that (2) itself holds for  $n = 3$ .

Our main result establishes (2) for any non-singular form with  $n \geq 10$ , and, moreover, applies to appropriate inhomogeneous polynomials.

**Theorem 2.** Let  $F(\mathbf{x}) = F(x_1, \dots, x_n) \in \mathbb{Z}[x_1, \dots, x_n]$ , where  $n \geq 5$ , and write  $F_0(\mathbf{x})$  for the homogeneous part of  $F$  of maximal degree. Suppose that  $F_0$  is non-singular, with degree at least 3. Then

$$N(F, X) \ll_F X^{n-3+15/(n+5)}.$$

This result improves on Theorem 1 as soon as  $n > 7$ . It is of course easy to see that (2) holds if  $F_0$  is a non-singular quadratic form, provided that  $n \geq 3$ . However a non-singular quadratic form in 3 or more variables either vanishes only at  $\mathbf{x} = 0$ , or satisfies  $N(F, X) \gg_F X^{n-2}$ . Thus Theorem 2 itself cannot hold for quadratic polynomials in general.

It is reasonable to expect that Theorem 2 should hold with exponent  $n - 3 + \varepsilon$ , but some restriction on  $n$  will always be necessary, as the example  $F = x_1^3 + x_2^3 + x_3^3 + x_4^3$  shows. This is non-singular, but has  $N(F, X) \gg X^2$ , in view of the trivial solutions  $(a, -a, b, -b)$ . Manin has formulated some very general conjectures covering such situations, see Franke *et al* [3], for example.

Our bounds for  $N(F, X)$  come from estimating the number of solutions of a congruence:

$$N(F, X, m) = \#\{\mathbf{x} \in \mathbb{Z}^n : F(\mathbf{x}) \equiv 0 \pmod{m}, |\mathbf{x}| \leq X\}.$$

One trivially has  $N(F, X) \leq N(F, X, m)$ , and we use two different methods to bound

The method of Cohen [2] does indeed do this, but the result already cited appears to be the best that one can achieve by this approach. One can also attempt to tackle the problem via the circle method. Thus, for example, the work of Birch [1] shows that  $N(F, X) \ll_F X^{n-d}$  for non-singular  $F$ , as soon as  $n > (d-1)2^d$ . However in the present paper the emphasis is on results that are independent of the degree  $d$  of  $F$ , whereas Birch's result only takes affect when  $n$  is sufficiently large in comparison to  $d$ .

For the proof of Theorem 1 we take  $m$  to be prime, and apply Deligne's estimate to bound  $N(F, X, m)$ . For Theorem 2 we shall take  $m$  to be composite and use the ' $q$ -analogue' of van der Corput's method. The interested reader is referred to the work of Heath-Brown [6], and of Graham and Ringrose [5] for other examples of this technique. We use the method in an  $n$ -dimensional setting, where ' $t$ -analogues' of van der Corput's method are extraordinarily cumbersome, to the extent that there has been no successful work for  $n \geq 3$ . However for our  $q$ -analogue we are able to carry out the van der Corput AB process in full. It is possible therefore that this work could provide a model for future investigations into multi-dimensional exponential sums.

During the course of the proof of Theorem 2 we shall require an estimate for the number of solutions, suitably weighted, of a polynomial congruence in several variables. When the homogeneous part of the polynomial is non-singular, such an estimate follows naturally from the work of Deligne. However we shall be concerned with the general case. Here Hooley [8] has investigated the situation for unweighted solutions, and our situation may be viewed as a generalization of his. It seems to us that the result merits formal statement as a theorem.

**Theorem 3.** *Let  $W(\mathbf{x})$  be an infinitely differentiable function, supported in a cube of side  $2L$ , and let  $f(\mathbf{x}) \in \mathbb{Z}[\mathbf{x}]$  be an arbitrary polynomial of degree  $d \geq 2$  in  $n$  variables. Write  $f_0$  for the homogeneous part of  $f$  of degree  $d$ , and define  $s = s(f)$  to be the dimension of the projective algebraic set defined over the field of  $q$  elements by  $\nabla f_0 = 0$ . Here  $q$  is a prime greater than  $d$ . Then for any positive  $N \ll q$  we have*

$$\sum_{\mathbf{x} \in \mathbb{Z}^n, q|f(\mathbf{x})} W\left(\frac{1}{N}\mathbf{x}\right) = q^{-1} \sum_{\mathbf{y} \in \mathbb{Z}^n} W\left(\frac{1}{N}\mathbf{y}\right) + O_{n,d,L}(D_{n+1} N^{s+1} q^{(n-s-1)/2}),$$

where  $D_k$  is the maximum of all  $k$ -th order partial derivatives of  $W(\mathbf{x})$ , taken over all  $\mathbf{x} \in \mathbb{R}^n$ .

Notice that  $s$  lies in the range  $-1 \leq s \leq n-1$ , and that  $s = -1$  means that  $f_0$  is non-singular. Taking  $N = q$  we essentially recover Hooley's result.

## 2. Proof of Theorem 1

To prove Theorem 1 we shall consider  $N(F, X, p)$  where  $p \geq X$  is prime. It is convenient to introduce the weight function

$$w(\mathbf{z}) = \prod_{r=1}^n \left( \frac{\sin \pi z_r}{\pi z_r} \right)^2,$$

whose Fourier transform is

$$\hat{w}(\mathbf{y}) = \prod_{r=1}^n (\max\{1 - |y_r|, 0\}).$$

We then have

$$N(F, X) \leq N(F, X, p) \ll \sum_{\mathbf{x} \in \mathbb{Z}^n, p|F(\mathbf{x})} w\left(\frac{1}{2X}\mathbf{x}\right).$$

Using the Poisson summation formula we see that the sum on the right is

$$\begin{aligned} & \sum_{\mathbf{y}(\bmod p), p|F(\mathbf{y})} \sum_{\mathbf{z} \in \mathbb{Z}^n} w\left(\frac{1}{2X}(\mathbf{y} + p\mathbf{z})\right) \\ &= \sum_{\mathbf{y}(\bmod p), p|F(\mathbf{y})} \sum_{\mathbf{u} \in \mathbb{Z}^n} \left(\frac{2X}{p}\right)^n e_p(\mathbf{u}, \mathbf{y}) \hat{w}\left(\frac{2X}{p}\mathbf{u}\right), \\ &= \left(\frac{2X}{p}\right)^n \sum_{\mathbf{u} \in \mathbb{Z}^n} \hat{w}\left(\frac{2X}{p}\mathbf{u}\right) S_p(\mathbf{u}), \end{aligned}$$

where  $e_p(x)$  is  $\exp(2\pi i x/p)$  as usual and

$$S_p(\mathbf{u}) = \sum_{\mathbf{y}(\bmod p), p|F(\mathbf{y})} e_p(\mathbf{u}, \mathbf{y}).$$

We therefore conclude that

$$N(F, X) \ll \left(\frac{X}{p}\right)^n \sum_{|\mathbf{u}| \ll p/X} |S_p(\mathbf{u})|. \quad (3)$$

When  $p|\mathbf{u}$  we have

$$S_p(\mathbf{u}) = \#\{\mathbf{y}(\bmod p): p|F(\mathbf{y})\} \ll p^{n-1}.$$

We therefore proceed to examine the case  $p \nmid \mathbf{u}$ . Since  $F$  is homogeneous we have

$$\begin{aligned} (p-1)S_p(\mathbf{u}) &= \sum_{a=1}^{p-1} \sum_{a\mathbf{y}(\bmod p), p|F(a\mathbf{y})} e_p(a\mathbf{u}, \mathbf{y}) \\ &= \sum_{a=1}^{p-1} \sum_{\mathbf{y}(\bmod p), p|F(\mathbf{y})} e_p(a\mathbf{u}, \mathbf{y}) \\ &= \sum_{\mathbf{y}(\bmod p), p|F(\mathbf{y})} \sum_{a=1}^{p-1} e_p(a\mathbf{u}, \mathbf{y}) \\ &= p\#\{\mathbf{y}(\bmod p): p|F(\mathbf{y}), p|\mathbf{u}, \mathbf{y}\} - \#\{\mathbf{y}(\bmod p): p|F(\mathbf{y})\}. \end{aligned}$$

Now, according to Deligne's estimate we have

$$\#\{\mathbf{y}(\bmod p): p|F(\mathbf{y}), p|\mathbf{u}, \mathbf{y}\} = p^{n-2} + O_{n,d}(p^{(n-1)/2}), \quad (4)$$

providing that  $F(\mathbf{y}) = \mathbf{u} \cdot \mathbf{y} = 0$  defines a non-singular absolutely irreducible variety of dimension  $n - 3$  over the field of  $p$  elements. Since  $F$  is non-singular there is a form  $G(\mathbf{u})$  with integer coefficients, depending only on  $F$ , such that (4) holds whenever  $p \nmid G(\mathbf{u})$ . The form  $G$  will be absolutely irreducible and will have degree 2 or more. The existence of such a form follows from the argument of Heath-Brown [7; Lemma 6], where forms  $F$  of degree 3 were considered. Since  $F$  is non-singular Deligne's estimate also gives

$$\#\{\mathbf{y}(\bmod p) : p \mid F(\mathbf{y})\} = p^{n-1} + O_{n,d}(p^{n/2})$$

for all sufficiently large primes  $p$ . It therefore follows that

$$\begin{aligned} S_p(\mathbf{u}) &\ll_F \frac{1}{p-1} \{p(p^{n-2} + O_{n,d}(p^{(n-1)/2})) - (p^{n-1} + O_{n,d}(p^{n/2}))\} \\ &\ll_F p^{(n-1)/2} \end{aligned}$$

whenever  $p \nmid G(\mathbf{u})$ . For the remaining values of  $\mathbf{u}$  we have

$$S_p(\mathbf{u}) = p^{-1} \sum_{a(\bmod p)} \sum_{\mathbf{y}(\bmod p)} e_p(aF(\mathbf{y}) + \mathbf{u} \cdot \mathbf{y}).$$

For  $p \nmid a$  the sum over  $\mathbf{y}$  is  $O_{n,d}(p^{n/2})$  by Deligne's bound for exponential sums, since  $F$  has degree at least 2. In the remaining case  $p \mid a$  the sum is zero, since we are assuming that  $p \nmid \mathbf{u}$ . Thus

$$S_p(\mathbf{u}) \ll_{n,d} p^{n/2}$$

whenever  $p \nmid \mathbf{u}$ .

We now insert our bounds for  $S_p(\mathbf{u})$  into the estimate (3) to obtain

$$N(F, X) \ll_F X^n p^{-1} + p^{(n-1)/2} + X^n p^{-n/2} \#\{|\mathbf{u}| \ll p/X : p \mid G(\mathbf{u})\}.$$

providing that  $p \geq X$ . Finally we average this result over all primes  $p$  in the interval  $[P, 2P]$ , so that

$$N(F, X) \ll_F X^n P^{-1} + P^{(n-1)/2} + X^n P^{-n/2} M, \quad (5)$$

where

$$\begin{aligned} M &= \frac{\log P}{P} \sum_p \#\{|\mathbf{u}| \ll P/X : p \mid G(\mathbf{u})\} \\ &\ll \frac{\log P}{P} \sum_{|\mathbf{u}| \ll P/X} \#\{p \mid G(\mathbf{u}) : P \leq p \leq 2P\}. \end{aligned}$$

However

$$G(\mathbf{u}) \ll |\mathbf{u}|^D \ll P^D,$$

where  $D$  is the degree of  $G$ , and hence a non-zero value of  $G(\mathbf{u})$  can have at most  $D < 1$  prime factors  $p \geq P$ . It follows that

$$M \ll \left(\frac{P}{X}\right)^n \frac{\log P}{P} + \#\{\mathbf{u} \ll P/X : G(\mathbf{u}) = 0\}$$

$$\ll \left(\frac{P}{X}\right)^n \frac{\log P}{P} + \left(\frac{P}{X}\right)^{n-3/2} \log P$$

on applying the bound (1) to the form  $G$ . In view of (5) we now have

$$N(F, X) \ll_f X^n P^{-1} + P^{(n-1)/2} + X^n P^{-n/2} \left( \left(\frac{P}{X}\right)^n \frac{\log P}{P} + \left(\frac{P}{X}\right)^{n-3/2} \log P \right).$$

We choose  $P = X^{2n/(n+1)} \geq X$ , which makes the first two terms above equal, and Theorem 1 follows.

### 3. Proof of Theorem 2: Preliminary manipulations

Deligne's estimate for 'complete' exponential sums, as used in the proof of Theorem 1, is essentially best possible, and the problem of improving on the corresponding estimate for 'incomplete' sums is notoriously difficult, even for sums in one variable. Thus one has

$$\sum_{1 \leq n \leq N} e_p(f(n)) \ll_f N p^{-1/2} + p^{1/2} \log p$$

for any non-constant integer polynomial  $f$ , and one suspects that better bounds should be possible when  $N$  is a suitable power of  $p$ . However in general nothing can be proved. To circumvent this difficulty we shall use composite moduli, where the 'q-analogue' of van der Corput's method is available. We begin by taking two primes  $p$  and  $q$ , with

$$p < X < q, \tag{6}$$

and considering  $N(F, X, pq)$ . It is convenient to use an infinitely differentiable weight with compact support, and we shall therefore define

$$\omega_0(t) = \begin{cases} \exp((1-t^2)^{-1}), & |t| < 1, \\ 0, & |t| \geq 1, \end{cases}$$

and

$$\omega(t) = \prod_{r=1}^n \omega_0(t_r).$$

Then  $\omega$  is infinitely differentiable, and its Fourier transform satisfies

$$\hat{\omega}(t) \ll_A |t|^{-A}, \quad |t| \geq 1, \tag{7}$$

for any  $A > 0$ . We now begin our analysis with the inequalities

$$N(F, X) \leq N(F, X, pq)$$



$$\begin{aligned}
&\ll \sum_{\mathbf{x} \in \mathbb{Z}^n, p|F(\mathbf{x})} \omega\left(\frac{1}{2X}\mathbf{x}\right) \\
&= S + K \#\{\mathbf{y}(\bmod p) : p|F(\mathbf{y})\} \\
&\ll S + Kp^{n-1},
\end{aligned} \tag{8}$$

where  $K$  is a parameter to be chosen and

$$\begin{aligned}
S &= \sum_{\mathbf{y}(\bmod p), p|F(\mathbf{y})} \left\{ \sum_{\mathbf{x} \equiv \mathbf{y}(\bmod p), q|F(\mathbf{x})} \omega\left(\frac{1}{2X}\mathbf{x}\right) - K \right\} \\
&\ll p^{(n-1)/2} \left\{ \sum_{\mathbf{y}(\bmod p), p|F(\mathbf{y})} \left| \sum_{\mathbf{x} \equiv \mathbf{y}(\bmod p), q|F(\mathbf{x})} \omega\left(\frac{1}{2X}\mathbf{x}\right) - K \right|^2 \right\}^{1/2} \\
&= p^{(n-1)/2} \Sigma^{1/2},
\end{aligned} \tag{9}$$

say. Here we have used Cauchy's inequality, together with the observation that there are  $O(p^{n-1})$  solutions of  $p|F(\mathbf{y})$  modulo  $p$ .

At this point, somewhat surprisingly, we include a number of extra terms in  $\Sigma$ . This, however, has the desirable effect of producing a sum in which there are only congruences modulo  $q$ . We have

$$\Sigma \leq \sum_{\mathbf{y}(\bmod p)} \sum_{a(\bmod q)} \left| \sum_{\substack{\mathbf{x} \equiv \mathbf{y}(\bmod p) \\ F(\mathbf{x}) \equiv a(\bmod q)}} \omega\left(\frac{1}{2X}\mathbf{x}\right) - K \right|^2. \tag{10}$$

When the sum on the right is expanded there are cross terms

$$K \sum_{\mathbf{y}(\bmod p)} \sum_{a(\bmod q)} \sum_{\substack{\mathbf{x} \equiv \mathbf{y}(\bmod p) \\ F(\mathbf{x}) \equiv a(\bmod q)}} \omega\left(\frac{1}{2X}\mathbf{x}\right) = K \sum_{\mathbf{x} \in \mathbb{Z}^n} \omega\left(\frac{1}{2X}\mathbf{x}\right).$$

Hence if we choose

$$K = p^{-n} q^{-1} \sum_{\mathbf{x} \in \mathbb{Z}^n} \omega\left(\frac{1}{2X}\mathbf{x}\right) \tag{11}$$

then (10) yields

$$\Sigma \leq \sum_{\mathbf{x} \in \mathbb{Z}^n} \omega\left(\frac{1}{2X}\mathbf{x}\right) \sum_{\substack{\mathbf{x}' \equiv \mathbf{x}(\bmod p), \\ F(\mathbf{x}') \equiv F(\mathbf{x})(\bmod q)}} \omega\left(\frac{1}{2X}\mathbf{x}'\right) - p^n q K^2.$$

On writing  $\mathbf{x}' = \mathbf{x} + p\mathbf{y}$  and

$$F(\mathbf{x}; \mathbf{y}) = F(\mathbf{x} + p\mathbf{y}) - F(\mathbf{x}), \quad W(\mathbf{x}; \mathbf{y}) = \omega\left(\frac{1}{2X}\mathbf{x}\right) \omega\left(\frac{1}{2X}(\mathbf{x} + p\mathbf{y})\right), \tag{12}$$

we find that

$$\Sigma \leq \sum_{\mathbf{y}} \sum_{q|F(\mathbf{x}; \mathbf{y})} W(\mathbf{x}; \mathbf{y}) - p^n q K^2. \tag{13}$$

The expected value of the sum over  $\mathbf{x}$  is

$$q^{-1} \sum_{\mathbf{x}} W(\mathbf{x}; \mathbf{y}) = K(\mathbf{y}), \quad (14)$$

say. By the Poisson summation formula we have

$$\begin{aligned} \sum_{\mathbf{y}} K(\mathbf{y}) &= q^{-1} \sum_{\mathbf{x}} \omega\left(\frac{1}{2X} \mathbf{x}\right) \sum_{\mathbf{y}} \omega\left(\frac{1}{2X} (\mathbf{x} + p\mathbf{y})\right) \\ &= q^{-1} \sum_{\mathbf{x}} \omega\left(\frac{1}{2X} \mathbf{x}\right) \left(\frac{2X}{p}\right)^n \sum_{\mathbf{z}} e_p(\mathbf{z}, \mathbf{x}) \hat{\omega}\left(\frac{2X}{p} \mathbf{z}\right). \end{aligned}$$

However, in view of (6) and (7), the terms with  $\mathbf{z} \neq 0$  contribute

$$O_A(q^{-1} X^n (X/p)^{n-A}).$$

Hence

$$\begin{aligned} \sum_{\mathbf{y}} K(\mathbf{y}) &= q^{-1} \sum_{\mathbf{x}} \omega\left(\frac{1}{2X} \mathbf{x}\right) \left(\frac{2X}{p}\right)^n \hat{\omega}(0) + O_A(q^{-1} X^n (X/p)^{n-A}) \\ &= (2X)^n K \hat{\omega}(0) + O_A(q^{-1} X^n (X/p)^{n-A}), \end{aligned}$$

by the definition (11). However, a second application of the Poisson summation formula yields

$$\begin{aligned} K &= p^{-n} q^{-1} \sum_{\mathbf{x}} \omega\left(\frac{1}{2X} \mathbf{x}\right) \\ &= p^{-n} q^{-1} (2X)^n \sum_{\mathbf{z}} \hat{\omega}(2X\mathbf{z}) \\ &= p^{-n} q^{-1} (2X)^n \hat{\omega}(0) + O_A(p^{-n} q^{-1} X^{n-A}). \end{aligned}$$

We therefore see that

$$\begin{aligned} \sum_{\mathbf{y}} K(\mathbf{y}) &= p^n q K^2 + O_A\left(\frac{X^n}{q} \left(\frac{X}{p}\right)^{n-A}\right) + O_A\left(\frac{X^n}{p^n q} X^{n-A}\right) + O_A\left(\frac{X^{2(n-A)}}{p^n q}\right) \\ &= p^n q K^2 + O_A\left(\frac{X^n}{q} \left(\frac{X}{p}\right)^{n-A}\right), \end{aligned}$$

so that (13) can be rewritten in the form

$$\Sigma \leq \sum_{\mathbf{y}} \left\{ \sum_{q|F(\mathbf{x}; \mathbf{y})} W(\mathbf{x}; \mathbf{y}) - K(\mathbf{y}) \right\} + O_A\left(\frac{X^n}{q} \left(\frac{X}{p}\right)^{n-A}\right). \quad (15)$$

We now combine the estimates (8), (9), (11) and (15) in the following lemma.

*Lemma 1. If  $p < X < q$  we have*

$$N(F, X) \ll \frac{X^n}{pq} + p^{(n-1)/2} \left\{ \sum_{\mathbf{y} \in \mathbb{Z}^n} |\Delta(\mathbf{y})| \right\}^{1/2} + \left\{ \frac{(pX)^n}{pq} \left(\frac{X}{p}\right)^{n-A} \right\}^{1/2},$$

for any  $A > 0$ , where

$$\Delta(\mathbf{y}) = \sum_{\mathbf{x} \in \mathbb{Z}^n, q|F(\mathbf{x}; \mathbf{y})} W(\mathbf{x}; \mathbf{y}) - K(\mathbf{y}).$$

#### 4. Proof of Theorem 2

We can now use Theorem 3, which we shall prove later, to estimate  $\Delta(\mathbf{y})$ . We take  $N = X$  and

$$W\left(\frac{1}{N}\mathbf{x}\right) = W(\mathbf{x}; \mathbf{y}) = \omega\left(\frac{1}{2X}\mathbf{x}\right)\omega\left(\frac{1}{2X}(\mathbf{x} + p\mathbf{y})\right),$$

so that  $W(\mathbf{x})$  is supported on a cube of side  $L = 4$ . With this choice of  $W(\mathbf{x})$  we have  $D_k \ll_k 1$ . According to the definition (14) we see that the main term in Theorem 3 is just  $K(\mathbf{y})$ , while the error term becomes  $O(X^{s+1}q^{(n-s-1)/2})$ , where  $s = s(F(\mathbf{x}; \mathbf{y})) = s(\mathbf{y})$ , say. With this notation it follows that

$$\Delta(\mathbf{y}) \ll q^{n/2}(Xq^{-1/2})^{1+s(\mathbf{y})}.$$

We now insert the above bound into Lemma 1. In doing so we observe that, according to the definition (12),  $W(\mathbf{x}; \mathbf{y})$  vanishes whenever  $|\mathbf{y}| \gg X/p$ . We therefore obtain

$$\begin{aligned} N(F, X) &\ll \frac{X^n}{pq} + p^{(n-1)/2}q^{n/4} \left\{ \sum_{|\mathbf{y}| \ll X/p} (Xq^{-1/2})^{1+s(\mathbf{y})} \right\}^{1/2} \\ &\quad + \left\{ \frac{(pX)^n}{pq} \left( \frac{X}{p} \right)^{n-4} \right\}^{1/2}. \end{aligned} \quad (16)$$

It remains to consider how often each value of  $s(\mathbf{y})$  can arise. The homogeneous part of maximal degree in

$$f(\mathbf{x}) = F(\mathbf{x}; \mathbf{y}) = F(\mathbf{x} + p\mathbf{y}) - F(\mathbf{x})$$

will be  $p\mathbf{y} \cdot \nabla F_0(\mathbf{x})$ , unless this happens to vanish identically in  $\mathbf{x}$ . We now use the following lemmas, which we will prove at the end of this section.

*Lemma 2.* Let  $f(\mathbf{x})$  be a non-singular form of degree  $d$  in  $n$  variables, defined over the field of  $q$  elements, where  $q > d$  is a prime. Let  $S_{\mathbf{y}} = S_{\mathbf{y}}(f)$  be the affine algebraic set

$$S_{\mathbf{y}} = \{\mathbf{x}; \mathbf{y} \cdot \nabla^2 f(\mathbf{x}) = 0\},$$

and let

$$T_s = \{\mathbf{y}; \dim(S_{\mathbf{y}}) \geq s\}.$$

Then  $T_s$  is a projective affine set of dimension at most  $n - s$ , defined by  $O_{n,d}(1)$  equations, each of degree  $O_{n,d}(1)$ .

*Lemma 3.* Let  $q$  be a prime, and let  $G_1, \dots, G_s$  be polynomials in  $n$  variables, defined

over the field of  $q$  elements, and producing an affine algebraic set all of whose components have dimension at most  $r \geq 0$ . Suppose further that the degrees of the polynomials  $G_i$  are all at most  $D$ . Then for any  $B$  with  $1 \leq B \ll q$  the number of integer solutions of the simultaneous congruences

$$G_i(\mathbf{x}) \equiv 0 \pmod{q}, \quad (1 \leq i \leq s)$$

in the region  $|\mathbf{x}| \leq B$  is  $O_{s,n,D}(B^r)$ .

Taking  $f = F_0$  we see that if  $\mathbf{y} \cdot \nabla F_0(\mathbf{x})$  vanishes identically in  $\mathbf{x}$ , then  $\mathbf{y} \in T_n$ , which has dimension zero, by Lemma 2. Hence there are only  $O_{n,d}(1)$  possible values of  $\mathbf{y}$  in the range  $|\mathbf{y}| \ll X/p$ . Here we use the observation that  $X/p < q$ . In the remaining cases Lemmas 2 and 3 show that there are  $O_{n,d}((X/p)^{n-s-1})$  values of  $\mathbf{y}$  with  $s(\mathbf{y}) = s$ . It therefore follows that

$$\begin{aligned} \sum_{|\mathbf{y}| \ll X/p} (Xq^{-1/2})^{1+s(\mathbf{y})} &\ll \sum_{s=-1}^{n-1} (Xq^{-1/2})^{1+s} (Xp^{-1})^{n-s-1} \\ &\ll (Xp^{-1})^n + (Xq^{-1/2})^n. \end{aligned}$$

The estimate (16) now yields

$$\begin{aligned} N(F, X) &\ll \frac{X^n}{pq} + p^{(n-1)/2} q^{n/4} \{ (Xp^{-1})^n + (Xq^{-1/2})^n \}^{1/2} \\ &\quad + \left\{ \frac{(pX)^n}{pq} \left( \frac{X}{p} \right)^{n-A} \right\}^{1/2}, \end{aligned}$$

for any  $A > 0$ , subject to the conditions  $p < X < q$ . We therefore choose primes  $p$  and  $q$  for which

$$X^{n/(n+5)} \ll p \ll X^{n/(n+5)}, \quad X^{2n/(n+5)} \ll q \ll X^{n/(n+5)}.$$

This is an admissible choice providing that  $n \geq 5$ , and yields

$$N(F, X) \ll X^{n-3+15/(n+5)},$$

providing that  $A$  is chosen large enough. This completes the proof of Theorem 2.

It remains to establish Lemmas 2 and 3. The result of Lemma 2 may be viewed as an extension of Lemma 2 of the author's work [7]. However the proof given there is not strictly correct. (The set  $\pi^{-1}(\mathbf{x})$  is  $\{(\mathbf{x}, \mathbf{y}) \in V : \mathbf{B}(\mathbf{x}, \mathbf{y}) = \mathbf{0}\}$  rather than  $\{(\mathbf{x}, \mathbf{y}) : \mathbf{B}(\mathbf{x}, \mathbf{y}) = \mathbf{0}\}$ .) We therefore use a slightly different approach.

We begin by showing that  $T_s$  is an algebraic set. To do this we take a generic linear space  $L$  of dimension  $n-s$ . Then  $T_s$  consists of those points  $\mathbf{y}$  for which  $S_{\mathbf{y}} \cap L$  is non-empty. However  $S_{\mathbf{y}} \cap L$  is given by  $n+s$  equations in  $\mathbf{x}$ , the first  $n$  of which are  $\mathbf{y} \cdot \nabla^2 f(\mathbf{x}) = \mathbf{0}$ , and the remaining  $s$  of which are the linear equations which specify that  $\mathbf{x} \in L$ . It follows from elimination theory that the condition for  $S_{\mathbf{y}} \cap L$  to be non-empty is the simultaneous vanishing of  $O_{n,d}(1)$  homogeneous polynomials of degrees  $O_{n,d}(1)$  in  $\mathbf{y}$ .

We have now to consider the dimension of  $T_s$ . We shall write  $S$  for the set

$$S = \{(\mathbf{x}, \mathbf{y}) : \mathbf{y} \cdot \nabla^2 f(\mathbf{x}) = \mathbf{0}\},$$

which we shall regard as a subset of  $2n$ -dimensional affine space. We begin by observing that  $\dim(S) \leq n$ , for otherwise  $S$  would have non-trivial points in common with the diagonal  $\{(\mathbf{x}, \mathbf{x})\}$ , which itself has dimension  $n$ . Any such common point would produce a non-zero solution  $\mathbf{x}$  of

$$(d-1)\nabla f(\mathbf{x}) = \mathbf{x} \cdot \nabla^2 f(\mathbf{x}) = 0,$$

contradicting the non-singularity assumption.

We now prove that  $\dim(T_s) \leq n-s$ . Let  $U$  be an irreducible component of  $T_s$ , and take  $\mathbf{P}$  to be a generic point of  $U$ . We write

$$W = \{(\mathbf{x}, \mathbf{y}) : \mathbf{y} \cdot \nabla^2 f(\mathbf{x}) = 0, \mathbf{y} \in U\},$$

and decompose  $W$  into irreducible components  $W_1 \cup \dots \cup W_r$ . With this notation it follows that

$$S_{\mathbf{P}} = \{\mathbf{x} : (\mathbf{x}, \mathbf{P}) \in W\} = \cup_{i=1}^r \{\mathbf{x} : (\mathbf{x}, \mathbf{P}) \in W_i\}.$$

Since  $\mathbf{P} \in T_s$ , at least one of the components of  $S_{\mathbf{P}}$  has dimension  $s$  or more. Consequently there is at least one index  $i$  for which

$$\dim(\{\mathbf{x} : (\mathbf{x}, \mathbf{P}) \in W_i\}) \geq s.$$

For ease of notation we take  $i = 1$ . We now consider the mapping

$$\pi : W_1 \rightarrow U$$

given by  $\pi(\mathbf{x}, \mathbf{y}) = \mathbf{y}$ . This is a regular mapping between irreducible varieties. Moreover there is at least one point of the form  $(\mathbf{x}, \mathbf{P})$  in  $W_1$ , so that the image of  $\pi$  contains the point  $\mathbf{P}$ . However the image of  $\pi$  will be a closed subset of  $U$ , and  $\mathbf{P}$  is generic on  $U$ , so that  $\pi$  must in fact be onto. It now follows by the theorem on the dimension of fibres (Shafarevich [9; page 60] for example) that

$$\dim(\pi^{-1}(\mathbf{P})) = \dim(W_1) - \dim(U).$$

We therefore deduce that

$$\begin{aligned} \dim(U) &= \dim(W_1) - \dim(\pi^{-1}(\mathbf{P})) \\ &\leq \dim(S) - \dim(\{\mathbf{x} : (\mathbf{x}, \mathbf{P}) \in W_i\}) \\ &\leq n-s. \end{aligned}$$

as required.

Lemma 3 is related to Lemma 2 of Hooley [8], and to the principle described by the author [7; page 229]. Indeed Hooley's result is essentially the case  $B = q$ . For the proof we first remark that the algebraic set

$$G : G_1 = \dots = G_s = 0$$

has  $O_{s,n,D}(1)$  components of dimension  $r$ . To show this we intersect  $G$  with a generic linear space  $L$  of dimension  $n-r$ . This will produce a set of isolated points only, at least one for each component of  $G$  of dimension  $r$ . Moreover distinct components

will produce distinct points of intersection with  $L$ . The number of components is thus at most the number of points in  $G \cap L$ . Since  $L$  is given by  $r$  linear equations,  $G \cap L$  is given by  $O_{s,n,D}(1)$  equations of degrees  $O_{s,n,D}(1)$ . A straightforward application of elimination theory now shows that  $G \cap L$  has  $O_{s,n,D}(1)$  points, and the result follows.

We can now give the proof of the lemma. We shall use induction on  $n$ , the result being trivial for  $n = 0$ . For the general case we write  $x = (x, y)$ , where  $y$  is an  $n - 1$  dimensional vector. Since the algebraic set defined by  $G_1 = \dots = G_s = 0$  has  $O_{n,D,s}(1)$  components of dimension  $r$ , there are  $O_{n,D,s}(1)$  values of  $x$  for which the hyperplane  $x_1 = x$  can contain such a component. For the remaining values of  $x$  every component of the set

$$x_1 = x, G_1 = \dots = G_s = 0$$

has dimension at most  $r - 1$ . Applying the case  $n - 1$  of the lemma to each hyperplane section there will be  $O_{n,D,s}(1)$  contributions  $O_{n,D,s}(B^r)$  from values of  $x$  of the first type, and  $O(B)$  contributions  $O_{n,D,s}(B^{r-1})$  from values of  $x$  of the second type. The case  $n$  of the lemma then follows.

### 5. Proof of Theorem 3

Before beginning the proof we record the remark that

$$D_k \ll_{n,L} D_{k+1},$$

which follows from the fact  $W$  is supported on a cube of side  $2L$ . The theorem will be proved by induction on  $s$ , and our first task is to establish the base step  $s = -1$ . We commence by using the manipulations with which we began the proof of Theorem 1. We have

$$\begin{aligned} \sum_{x \in \mathbb{Z}^n, q|f(x)} W\left(\frac{1}{N}x\right) &= \sum_{z \pmod{q}, q|f(z)} \sum_{u \in \mathbb{Z}^n} W\left(\frac{1}{N}(z + qu)\right) \\ &= \sum_{z \pmod{q}, q|f(z)} \left(\frac{N}{q}\right)^n \sum_{v \in \mathbb{Z}^n} e_q(v \cdot z) \hat{W}\left(\frac{N}{q}v\right) \\ &= \left(\frac{N}{q}\right)^n \sum_{v \in \mathbb{Z}^n} \hat{W}\left(\frac{N}{q}v\right) \Sigma_q(v), \end{aligned}$$

where

$$\begin{aligned} \Sigma_q(v) &= \sum_{z \pmod{q}, q|f(z)} e_q(v \cdot z) \\ &= q^{-1} \sum_{a \pmod{q}} \sum_{z \pmod{q}} e_q(af(z) + v \cdot z). \end{aligned}$$

When  $q \nmid a$  the homogeneous part of  $af(z) + v \cdot z$  of maximal degree is just  $af_q(z)$ , since  $d \geq 2$ . However we are assuming that  $f_0$  is non-singular, whence Deligne's estimate shows that the  $z$  sum is  $O_{n,d}(q^{n/2})$ . When  $q|a$  the sum is  $q^n$  for  $q|v$ , and vanishes in

the remaining cases. It therefore follows that

$$\begin{aligned} \sum_{\mathbf{x} \in \mathbb{Z}^n, q|f(\mathbf{x})} W\left(\frac{1}{N}\mathbf{x}\right) &= \frac{N^n}{q} \sum_{\mathbf{w} \in \mathbb{Z}^n} \widehat{W}(N\mathbf{w}) + O_{n,d}\left(\frac{N^n}{q^{n/2}} \sum_{\mathbf{v} \in \mathbb{Z}^n} \left| \widehat{W}\left(\frac{N}{q}\mathbf{v}\right) \right| \right) \\ &= q^{-1} \sum_{\mathbf{y} \in \mathbb{Z}^n} W\left(\frac{1}{N}\mathbf{y}\right) + O_{n,d}\left(\frac{N^n}{q^{n/2}} \sum_{\mathbf{v} \in \mathbb{Z}^n} \left| \widehat{W}\left(\frac{N}{q}\mathbf{v}\right) \right| \right), \end{aligned}$$

on using the Poisson summation formula again.

On integrating by parts  $n+1$  times one finds that

$$\widehat{W}(\mathbf{t}) \ll_{n,L} D_{n+1} |\mathbf{t}|^{-n-1}$$

for  $|\mathbf{t}| \geq 1$ . For smaller  $\mathbf{t}$  one may use the trivial bound

$$\widehat{W}(\mathbf{t}) \ll_{n,L} D_0 \ll_{n,L} D_{n+1}.$$

These estimates show that

$$\sum_{\mathbf{v} \in \mathbb{Z}^n} \left| \widehat{W}\left(\frac{N}{q}\mathbf{v}\right) \right| \ll_{n,L} D_{n+1} \left(\frac{q}{N}\right)^n,$$

and hence

$$\sum_{\mathbf{x} \in \mathbb{Z}^n, q|f(\mathbf{x})} W\left(\frac{1}{N}\mathbf{x}\right) = q^{-1} \sum_{\mathbf{y} \in \mathbb{Z}^n} W\left(\frac{1}{N}\mathbf{y}\right) + O_{n,d,L}(D_{n+1} q^{n/2}),$$

as required.

For the induction step we shall count points of affine hyperplanes  $\mathbf{m} \cdot \mathbf{x} = c$ . For any non-zero vector  $\mathbf{m}$  we therefore define a form  $f_0^{(\mathbf{m})}$  corresponding to the intersection of  $f_0 = 0$  with  $\mathbf{m} \cdot \mathbf{x} = 0$ . For an explicit representation of  $f_0^{(\mathbf{m})}$  one can label the coordinates so that  $m_1 \neq 0$ , and set

$$f_0^{(\mathbf{m})}(x_2, \dots, x_n) = f_0\left(-\frac{m_2 x_2 + \dots + m_n x_n}{m_1}, x_2, \dots, x_n\right).$$

In the next section we shall prove the following lemma.

*Lemma 4. Suppose that  $s(f_0) \geq 0$ . Then there is a non-zero integer vector  $\mathbf{m}$ , with*

$$\mathbf{m} \ll_{d,n} 1,$$

*for which  $s(f_0^{(\mathbf{m})}) = s(f_0) - 1$ .*

In particular  $f_0^{(\mathbf{m})}$  is not identically zero. Clearly we can assume that  $\mathbf{m}$  is primitive, so that there is a unimodular  $n \times n$  matrix  $M$  all of whose entries are  $O_{n,d}(1)$  and such that the first column of  $(M^T)^{-1}$  is the vector  $\mathbf{m}$ . We now define a new polynomial and a new weight function by

$$f_M(\mathbf{x}) = f(M\mathbf{x}), \quad W_M(\mathbf{x}) = W(M\mathbf{x}).$$

Since the entries of  $M^{-1}$  are all  $O_{n,d}(1)$  it follows that  $W_M$  is supported on a cube of

side  $2L'$ , depending only on  $n, d$  and  $L$ . We now have

$$\begin{aligned} \sum_{\mathbf{x} \in \mathbb{Z}^n, q|f(\mathbf{x})} W\left(\frac{1}{N}\mathbf{x}\right) &= \sum_{\mathbf{x} \in \mathbb{Z}^n, q|f_M(\mathbf{x})} W_M\left(\frac{1}{N}\mathbf{x}\right) \\ &= \sum_{-NL' \leq x \leq NL'} \sum_{y \in \mathbb{Z}^{n-1}, q|f_M(x, y)} W_M\left(\frac{1}{N}(x, y)\right). \end{aligned} \quad (17)$$

Here we have used the fact that  $W_M(\frac{1}{N}(x, y)) = 0$  unless  $-NL' \leq x \leq NL'$ , since  $W_M$  is supported in a cube of side  $2L'$ . Now if  $g(y) = f_M(x, y)$  then  $g_0(y) = f_0(M(0, y))$ . Since  $M^T \mathbf{m} = (1, 0, \dots, 0)$  we have  $\mathbf{m} \cdot M\mathbf{x} = (M^T \mathbf{m}) \cdot \mathbf{x} = x_1$ . However  $f_0^{(\mathbf{m})}$ , which corresponds to substituting  $\mathbf{m} \cdot \mathbf{x} = 0$  in  $f_0$ , is equivalent, under linear substitution, to the form obtained by substituting  $\mathbf{m} \cdot M\mathbf{x} = x_1 = 0$  in  $f_0(M\mathbf{x})$ . It follows that  $f_0^{(\mathbf{m})}$  is equivalent to  $g_0(y)$ , so that  $s(g) = s - 1$ .

We are now ready to apply to induction assumption to the inner sum in (17). We have

$$\begin{aligned} \sum_{y \in \mathbb{Z}^{n-1}, q|f_M(x, y)} W_M\left(\frac{1}{N}(x, y)\right) \\ = q^{-1} \sum_{y \in \mathbb{Z}^{n-1}} W_M\left(\frac{1}{N}(x, y)\right) + O_{n-1, d, L}(D_n(M) N^s q^{(n-s-1)/2}), \end{aligned}$$

where  $D_n(M)$  is the maximum of all  $n$ -th order partial derivatives of  $W_M(x, y)$ . In view of the definition of  $W_M$  we have

$$D_n(M) \ll_{n, d, L} D_n,$$

whence, on summing over  $x$ , we see from (17) that

$$\begin{aligned} \sum_{\mathbf{x} \in \mathbb{Z}^n, q|f(\mathbf{x})} W\left(\frac{1}{N}\mathbf{x}\right) &= \sum_{-NL' \leq x \leq NL'} \sum_{y \in \mathbb{Z}^{n-1}, q|f_M(x, y)} W_M\left(\frac{1}{N}(x, y)\right) \\ &= \sum_{-NL' \leq x \leq NL'} q^{-1} \sum_{y \in \mathbb{Z}^{n-1}} W_M\left(\frac{1}{N}(x, y)\right) \\ &\quad + O_{n, d, L}(D_n N^{s+1} q^{(n-s-1)/2}). \end{aligned}$$

Since  $D_n \ll D_{n+1}$ , the error term is of the form required for Theorem 3. Moreover, since,  $W_M$  is supported on a cube of side  $2L'$ , the main term is just

$$\begin{aligned} q^{-1} \sum_{\mathbf{x} \in \mathbb{Z}^n} \sum_{y \in \mathbb{Z}^{n-1}} W_M\left(\frac{1}{N}(x, y)\right) &= q^{-1} \sum_{\mathbf{x} \in \mathbb{Z}^n} W_M\left(\frac{1}{N}\mathbf{x}\right) \\ &= q^{-1} \sum_{\mathbf{x} \in \mathbb{Z}^n} W\left(\frac{1}{N}\mathbf{x}\right), \end{aligned}$$

since  $M$  is a unimodular integer matrix. The main term is therefore also of the form



Lemma 4 will be deduced from Lemma 3, together with the following result.

*Lemma 5. Let  $q$  be a prime, and let  $g$  be a homogeneous form of degree  $d$  in  $n$  variables, defined over the field  $K$  of  $q$  elements. Suppose further that  $2 \leq d < q$ . For any non-zero vector  $\mathbf{m} \in K^n$  let  $g^{(\mathbf{m})}$  be the form in  $n-1$  variables obtained by substituting  $\mathbf{m} \cdot \mathbf{x} = 0$  in  $g(\mathbf{x})$ , and let  $s(g)$  and  $s(g^{(\mathbf{m})})$  be defined as in Theorem 3. Then  $s(g^{(\mathbf{m})}) \geq s(g) - 1$  for all non-zero  $\mathbf{m}$ . Moreover if  $s(g) \neq -1$ , there is a non-zero form  $G$  depending on  $g$ , such that the degree of  $G$  is bounded in terms of  $n$  and  $d$  alone, and such that  $s(g^{(\mathbf{m})}) = s(g) - 1$  whenever  $q \nmid G(\mathbf{m})$ .*

Lemma 4 clearly follows from Lemmas 3 and 5, since Lemma 3, with  $s = 1$ ,  $r = n - 1$  and  $D = O_{n,d}(1)$  will produce a non-zero  $\mathbf{m}$  with  $q \nmid G(\mathbf{m})$  and  $|\mathbf{m}| \leq B$ , as soon as  $B \gg_{n,d} 1$ .

Lemma 5 is essentially contained in § 2 of Hooley [8]. However, since the result we require is not explicitly stated by Hooley, we shall give an appropriate modification of Hooley's treatment here.

We begin our proof by making a linear change of variables so that the hyperplane  $\mathbf{m} \cdot \mathbf{x} = 0$  becomes  $x_1 = 0$ . The singular loci  $\mathcal{S}$  and  $\mathcal{S}^{(\mathbf{m})}$  of  $g$  and  $g^{(\mathbf{m})}$  are then given by the systems

$$\frac{\partial g}{\partial x_1} = \dots = \frac{\partial g}{\partial x_n} = 0$$

and

$$x_1 = \frac{\partial g}{\partial x_2} = \dots = \frac{\partial g}{\partial x_n} = 0$$

respectively. We shall write  $\mathcal{L}$  for the hyperplane  $x_1 = 0$  and  $\mathcal{S}_0$  for the set

$$\frac{\partial g}{\partial x_2} = \dots = \frac{\partial g}{\partial x_n} = 0.$$

Thus

$$\mathcal{S}^{(\mathbf{m})} \cap \left\{ \frac{\partial g}{\partial x_1} = 0 \right\} = \mathcal{S} \cap \mathcal{L},$$

whence

$$\begin{aligned} s(g^{(\mathbf{m})}) &= \dim(\mathcal{S}^{(\mathbf{m})}) \\ &\geq \dim\left(\mathcal{S}^{(\mathbf{m})} \cap \left\{ \frac{\partial g}{\partial x_1} = 0 \right\}\right) \\ &= \dim(\mathcal{S} \cap \mathcal{L}) \\ &\geq \dim(\mathcal{S}) - 1 \\ &= s(g) - 1. \end{aligned}$$

It follows that  $s(g^{(\mathbf{m})}) \geq s(g) - 1$ , as required. Moreover if  $s(g^{(\mathbf{m})}) \neq s(g) - 1$ , then either

$\dim(\mathcal{S} \cap \mathcal{L}) = \dim(\mathcal{S})$  or

$$\dim(\mathcal{S}^{(\mathbf{m})}) > \dim\left(\mathcal{S}^{(\mathbf{m})} \cap \left\{\frac{\partial g}{\partial x_1} = 0\right\}\right). \quad (18)$$

In the former case one sees that  $\mathcal{L}$  must contain an irreducible component of  $\mathcal{S}$  of maximal dimension. Reverting to our original coordinate system we therefore let  $\mathcal{M}_1$  be the set of non-zero vectors  $\mathbf{m}$  for which the hyperplane  $\mathbf{m} \cdot \mathbf{x} = 0$  contains an irreducible component of the singular locus  $\nabla g = 0$  of maximal dimension. The number  $D'$ , say of such components satisfies  $D' = O_{n,d}(1)$  (as in the proof of Lemma 3) and picking a point  $\mathbf{p}_i$  from each we deduce that

$$G_0(\mathbf{m}) = \prod_i \mathbf{p}_i \cdot \mathbf{m} = 0$$

whenever  $\mathbf{m} \in \mathcal{M}_1$ . It may happen that the form  $G_0$  is defined over a finite extension  $K'$  of  $K$ , in which case a suitable constant multiple of the form  $G_0$  will have a trace  $G_1$  which does not vanish identically. Then  $G_1$  is defined over  $K$ , has degree  $D'$ , and has the property that  $G_1(\mathbf{m}) = 0$  whenever  $\mathbf{m} \in \mathcal{M}_1$ .

We write  $\mathcal{M}_2$  for the set of non-zero vectors  $\mathbf{m}$  for which (18) holds. For such an  $\mathbf{m}$  there is a point  $\mathbf{p}$ , say, in  $\mathcal{S}^{(\mathbf{m})}$ , for which  $\partial g / \partial x_1 \neq 0$ . It follows that  $p_1 = 0$ , since  $\mathbf{p}$  is on  $\mathcal{L}$ , and that

$$\frac{\partial g}{\partial x_2} = \dots = \frac{\partial g}{\partial x_n} = 0$$

at  $\mathbf{p}$ . We therefore see that  $\mathbf{p} \cdot \nabla g(\mathbf{p}) = 0$ , whence  $g(\mathbf{p}) = 0$ . Moreover  $\nabla g(\mathbf{p})$  is proportional to the vector  $(1, 0, 0, \dots, 0)$ . On returning to our original coordinate system we see that every  $\mathbf{m} \in \mathcal{M}_2$  can be represented as  $\nabla g(\mathbf{p})$  for some  $\mathbf{p}$  on the hypersurface  $g = 0$ . Here  $\mathbf{p}$  may be defined over some finite extension of  $K$ .

We now show that there is a non-zero form  $G_2$  of degree  $D'' = O_{n,d}(1)$  such that  $G_2(\nabla g(\mathbf{x}))$  is identically divisible by  $g(\mathbf{x})$ . To do this we observe, via elimination theory, that for any  $\mathbf{y}$ , the equation  $\nabla g(\mathbf{x}) = \mathbf{y}$  is solvable with  $g(\mathbf{x}) = 0$ , if and only if  $\mathbf{y}$  satisfies one of a set of conditions

$$C_i: E_{1,i}(\mathbf{y}) = 0 \quad \text{and} \quad E_{2,i}(\mathbf{y}) \neq 0, \quad (1 \leq i \leq I).$$

Here  $E_{1,i}$  and  $E_{2,i}$  are forms of degrees  $O_{n,d}(1)$  and  $I = O_{n,d}(1)$ . If  $E_{1,i}$  were to vanish identically for every  $i$  then the image of  $g = 0$  under  $\nabla$  would be a non-empty Zariski open subset of an affine  $n$ -space, and would therefore have dimension  $n$ . Since  $g = 0$  only has dimension  $n - 1$ , we conclude that at least one of the forms  $E_{1,i}$  does not vanish. Taking  $G_2$  to be any such form we have  $G_2(\nabla g(\mathbf{x})) = 0$  whenever  $g(\mathbf{x}) = 0$ , whence  $g(\mathbf{x})$  divides  $G_2(\nabla g(\mathbf{x}))$  as required.

Finally we observe that  $G_2(\mathbf{m}) = 0$  whenever  $\mathbf{m} \in \mathcal{M}_2$ . Lemma 5 now follows on taking  $G = G_1 G_2$  and  $D = D' + D''$ .

## 7. Acknowledgement

It is a pleasure to record the support of the Isaac Newton Institute, Cambridge, where this work was carried out.

## References

- [1] Birch B J, Forms in many variables, *Proc. R. Soc. A* **265** (1961/62) 245–263
- [2] Cohen S D, The distribution of Galois groups and Hilbert's irreducibility theorem, *Proc. Lond. Math. Soc.* (3), **43** (1981) 227–250
- [3] Franke J, Manin Y I and Tschinkel Y, Rational points of bounded height on Fano varieties, *Invent. Math.* **95** (1989) 421–435
- [4] Fujiwara M, Upper bounds for the number of lattice points on hypersurfaces, *Number theory and combinatorics, Japan, 1984* (Singapore: World Scientific Publishing Co.) (1985)
- [5] Graham S W and Ringrose C J, Lower bounds for least quadratic non-residues, Analytic number theory, Allerton Park, II 1989 *Prog. Math.*, **85** (Boston: Birkhauser) (1990) 245–263
- [6] Heath-Brown D R, Hybrid bounds for  $L$ -functions: a  $q$ -analogue of van der Corput's method and a  $t$ -analogue of Burgess' method, *Recent progress in analytic number theory, Vol. I*, (London: Academic Press) (1981) 121–126
- [7] Heath-Brown D R, Cubic forms in ten variables, *Proc. Lond. Math. Soc.* (3), **47** (1983), 225–257
- [8] Hooley C, On the number of points on a complete intersection over a finite field, *J. Number Theory*, **38** (1991) 338–358
- [9] Shararevich I R, *Basic algebraic geometry*, (New York: Springer) (1977)



ANATOLI ANDRIANOV\*

Sonderforschungsbereich 170, "Geometrie und Analysis", Mathematisches Institut der Universität, Bunsenstr. 3-5, D-3400 Göttingen, Germany

\*Permanent address: St. Petersburg Branch of the Steklov Mathematical Institute, Fontanka 27, 191011 St. Petersburg, Russia

Dedicated to the memory of Professor K G Ramanathan

**Abstract.** Let  $q(X)$  be a quadratic form in an even number  $m$  of variables with coefficients in a Dedekind ring  $K$ . Let us assume that the sets

$$R(q, a) = \{N \in K^m; q(N) = a\}$$

of representations of elements  $a$  of  $K$  by the form  $q$  are finite. Then certain multiplicative relations are obtained by elementary means between the sets  $R(q, a)$  and  $R(q, ab)$ , where  $b$  is a product of prime elements  $\rho$  of  $K$  with finite coefficients  $K/\rho K$ . The relations imply similar multiplicative relations between the numbers of elements of the sets  $R(q, a)$ , which formerly could be obtained only in some special cases like the case when  $K = \mathbb{Z}$  is the ring of rational integers and only by means of the theory of Hecke operators on the spaces of theta-series. As an application, an almost elementary proof of the Siegel theorem on the mean number of representations of integers by integral positive quadratic forms of determinant 1 is given.

**Keywords.** Quadratic forms; multiplicative properties; rings of automorphs; Siegel theorem.

## 1. Introduction

We consider quadratic forms

$$q(X) = \sum_{1 \leq i \leq j \leq m} q_{ij} x_i x_j, \quad \text{where } X = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}, \quad (1.1)$$

with coefficients  $q_{ij}$  in a Dedekind ring  $K$  and we shall be interested in relations between the sets

$$R(q, a) = R_K(q, a) = \{N \in K^m; q(N) = a\} \quad (1.2)$$

of the representations (over  $K$ ) of various elements  $a$  of  $K$  by the form  $q$ . It turns out that under certain conditions on the form  $q$  and a principal prime ideal  $\rho K$  of  $K$  each representation  $N \in R(q, \rho a)$  can be factorized in the form

$$N = DN',$$

where  $m \times m$ -matrices  $D$  over  $K$  satisfy

$$q(DX) = \rho q'(X)$$

with suitable quadratic forms  $q'$  over  $K$ , and where  $N'$  belongs to  $R(q', a)$ . The explicit form of the factorization is given by Theorem 4.3. Consecutive application of the theorem to various principal prime ideals and their degrees reveals certain multiplicative properties of the sets  $R(q, a)$ , given by Corollaries 4.4 and 4.5. As to the quadratic forms under consideration, we assume that they are non-degenerate forms in an even number of variables. Those are the sufficient conditions to apply our main technical tool, the lifting theorem 2.1. We make also a natural assumption on finiteness of class number of the forms (see (2.20)). In addition, we assume in this paper that the forms are *finite*, i.e. each of the sets of representations  $R(q, a)$  is finite. This will allow us to simplify the formulas and their proofs. The multiplicative relations between the sets  $R(q, a)$  for finite forms imply similar relations between the numbers of their elements

$$r(q, a) = |R(q, a)|,$$

which are given by Theorem 5.1 and formulas (5.3). Formerly, similar general relations could be obtained only for certain specific rings like the ring  $\mathbb{Z}$  of rational integers, using the theory of Hecke operators on corresponding spaces of modular forms. Finally, by specializing the relations to the case of positive definite quadratic forms of determinant 1 over the ring  $\mathbb{Z}$  we obtain an "almost" elementary proof of the Siegel theorem on mean numbers of representations for these forms.

*Notation.* If  $A$  is a set, then  $A_n^m$  denotes the set of all  $m \times n$ -matrices with entries in  $A$ ,  $A^m = A_1^m$ . If  $M$  is a matrix, then  ${}^tM$  is the transposed matrix.  $1_n$  will denote the unit matrix of order  $n$  over the considered ring.

The letters  $\mathbb{Z}$  and  $\mathbb{C}$  are reserved for the ring of rational integers and the field of complex numbers respectively. The ground ring  $K$  is usually supposed to be a *Dedekind ring*, i.e. a commutative integral domain, where each ideal is a product of prime ideals.

## 2. Lifting of solutions of quadratic congruences

We associate to each quadratic form (1.1) over a ring  $K$  the *matrix* of the form given by

$$Q = Q(q) = (q_{ij}) + {}^t(q_{ij}). \quad (2.1)$$

This is a symmetric  $m \times m$ -matrix over  $K$  whose diagonal entries belong to  $2K$ . For brevity, such a matrix will be called an *even* matrix of order  $m$  over  $K$  and the set of all even matrices of order  $m$  over  $K$  will be denoted by

$$\mathbb{E}_m = \mathbb{E}_m(K). \quad (2.2)$$

It is clear that each matrix  $Q \in \mathbb{E}_m$  is the matrix of a quadratic form  $q$  in  $m$  variables over  $K$ , which satisfies

$$2q(X) = {}^tXQX. \quad (2.3)$$

The element

$$d(q) = \det Q, \quad (2.4)$$

where  $Q$  is the matrix of a form  $q$ , is called *the determinant of  $q$* . If  $q$  is a quadratic form in  $m$  variables and  $M \in K_n^m$  with some  $n = 1, 2, \dots$ , we set

$$q|M = (q|M)(X) = q(MX), \quad (2.5)$$

which is clearly a quadratic form in  $n$  variables over  $K$ . In this case, we say that  $M$  is a *representation* of the form  $q' = q|M$  by the form  $q$  (over  $K$ ), and we denote by

$$R(q, q') = R_K(q, q') = \{M \in K_n^m; q|M = q'\} \quad (2.6)$$

the set of all representations of  $q'$  by  $q$ , where equality of quadratic forms is understood coefficient-wise with respect to the "triangular" form (1.1). If  $n = 1$ , so that  $q'(x) = ax^2$ , it follows from the definitions that the set (2.6) coincides with the set (1.2) of representations of element  $a$  by  $q$ :

$$R(q, ax^2) = R(q, a). \quad (2.7)$$

It is clear that the matrix  $Q'$  of the form  $q' = q|M$  is given by

$$Q' = Q(q|M) = {}^tMQM \quad (2.8)$$

where  $Q$  is the matrix of  $q$ . It follows that

$$R(q, q') = R(Q, Q') = \{M \in K_n^m; {}^tMQM = Q'\}$$

provided that characteristic of  $K$  is not 2.

To study the multiplicative dependence of representations  $N \in R(q, \rho a)$  on  $\rho$  and  $a$ , we consider them as solutions of quadratic congruences

$$q(N) \equiv 0 \pmod{\rho}, \quad (2.9)$$

which, according to (2.7), can be written in the form

$$q|N \equiv 0 \pmod{\rho}, \quad (2.10)$$

and ask in how many ways it is possible to "lift" a solution  $N \in K^m$ , that is to find a matrix  $D \in K_m^m$  which satisfies the congruence

$$q|D \equiv 0 \pmod{\rho} \quad (2.11)$$

and divides  $N$  in the sense that  $N \in DK^m$ , where congruences for quadratic forms are understood coefficient-wise. The answer is essentially given by the following lifting theorem.

**Theorem 2.1.** *Let  $q(X)$  be a quadratic form in an even number  $m = 2k$  of variables over a Dedekind ring  $K$  and let  $\rho K$  be a non-trivial principal prime ideal of  $K$  with finite residue class ring  $K/\rho K$  of  $n(\rho)$  elements. Let us assume that*

$$d = \det q \neq 0 \text{ and } \rho \text{ does not divide } d.$$

Then, for each column  $N \in K^m$  satisfying the congruence (2.9), the number of matrices

$$D \in \Lambda d_k(\rho) \Lambda / \Lambda, \text{ where } \Lambda = \Lambda^m = GL_m(K) \text{ and } d_k(\rho) \equiv \begin{pmatrix} 1_k & 0 \\ 0 & \rho 1_k \end{pmatrix}$$

satisfying the congruence (2.11) and dividing  $N$  depends only on whether  $N$  is congruent to zero modulo  $\rho$  or not, and can be given by the following formula

$$\sum_{\substack{D \in \Lambda d_k(\rho) \Lambda / \Lambda, \\ q|D \equiv 0 \pmod{\rho}, D|N}} 1 = c_\rho(q)(1 + \delta(N/\rho)\varepsilon_\rho(q)n(\rho)^{k-1}), \quad (2.12)$$

where

$$c_\rho(q) = \begin{cases} \prod_{i=0}^{k-2} (1 + \varepsilon_\rho(q)n(\rho)^i), & \text{if } m = 2k > 2 \\ 1, & \text{if } m = 2, \end{cases} \quad (2.13)$$

$\varepsilon_\rho(q) = \pm 1$  is the sign of the form  $q$  modulo  $\rho$ , i.e. the sign of the non-degenerate quadratic space  $((K/\rho K)^m, q \bmod \rho)$  defined by the form  $q$  modulo  $\rho$  over the finite field  $K/\rho K$ ; in particular, if the characteristic of the field is not 2, then

$$\varepsilon_\rho(q) = \begin{cases} 1, & \text{if } (-1)^k \det q \text{ is a square modulo } \rho \\ -1, & \text{otherwise,} \end{cases}$$

and where, for a matrix  $M$  over the quotient field of  $K$ , we set

$$\delta(M) = \begin{cases} 1 & \text{if all the entries of } M \text{ belong to } K \\ 0 & \text{otherwise.} \end{cases}$$

*Proof.* The theorem is just a specialization of the Theorem 5.1 [1] to the case when  $m = 2k$ ,  $r = k$  and  $n = 1$ .  $\square$

Note that the factor  $c_\rho(q)$  in (2.12) satisfies

$$c_\rho(q) \neq 0 \text{ if and only if } m = 2 \text{ or } m > 2 \text{ and } \varepsilon_\rho(q) = 1. \quad (2.14)$$

Now we are going to transform the formula (2.12) for further use, but first we introduce some definitions and notations. For an element  $a \in K$  we shall call the number (finite or infinite)

$$n(a) = |K/aK|$$

the *norm* of  $a$ . An element  $\rho \in K$  will be called *prime*, if the ideal  $\rho K$  is non-trivial and prime. By

$$P = P(q) \quad (2.15)$$

will be denoted a set of representatives modulo invertible elements (units) of  $K$  of all prime elements  $\rho \in K$  of finite norm which do not divide the determinant  $d = \det q$  and satisfy the condition (2.14).

For each  $\rho \in P$  and each  $N \in K^m$  satisfying (2.9) or (2.10), we can write the formula



(2.12). Each matrix  $D$  appearing in the formula defines a quadratic form  $q'$  over  $K$  satisfying the relation

$$q|D = \rho q'. \quad (2.16)$$

If  $Q$  and  $Q'$  are the matrices of  $q$  and  $q'$  respectively and  $d$  and  $d'$  are their determinants, it follows from (2.8) and (2.16) that

$${}^tDQD = \rho Q' \quad (2.17)$$

and so

$$(\det D)^2 d = \rho^m d'.$$

Since  $D \in \Lambda d_k(\rho)\Lambda$ , it follows that  $\det D = \rho^k \eta$  and so  $d' = \eta^2 d$ , where  $\eta$  is invertible in  $K$ . Further, the replacement of  $D$  by another representative  $DU$  with  $U \in \Lambda$  in the same coset  $D\Lambda$  replaces the form  $q'$  by another form  $q'|U$  in the same class

$$\{q'\} = \{q'|U; U \in \Lambda\} \quad (2.18)$$

of equivalent quadratic forms in  $m$  variables over  $K$ , and each form of the class can be obtained in this way. Note that the determinants of all quadratic forms of the same class belong to the same coset modulo the group of squares of invertible elements of  $K$ . The coset is called the *determinant of the class*. The above considerations show that each quadratic form  $q'$  associated to a matrix  $D$  in (2.12) by the relation (2.16) belongs to a class of determinant  $d = \det q$  and can be replaced by an arbitrary representative of the class by replacing  $D$  in  $D\Lambda$ . Because  $n(\rho) < \infty$ , it follows that the set  $\Lambda d_k(\rho)\Lambda/\Lambda$  is finite (see, for example, [1], § 2), and so the forms  $q'$  related to various  $D$  in (2.12) belong to a finite set of the classes. But to apply the formulas, a stronger finiteness condition will be required. We shall assume that the form  $q$  satisfies the following finiteness condition: The set  $S(q)$  of all quadratic forms  $q'$  in  $m$  variables over  $K$  belonging to classes of the same determinant as that of  $\{q\}$  and satisfying a condition of the form

$$q|D = \mu q' \quad (2.19)$$

with  $D \in K_m^m$  and a non-zero  $\mu \in K$  is a finite union of different classes (2.18):

$$S(q) = \bigcup_{i=1}^h \{q_i\}. \quad (2.20)$$

We shall denote by

$$\langle q \rangle = (q_1, \dots, q_h) \quad (2.21)$$

a system of representatives of the classes. Without loss of generality, one can assume that

$$\det q_1 = \dots = \det q_h = \det q. \quad (2.22)$$

Let us return to the formula (2.12). If for  $D \in K_m^m$  and  $D' = DU$  with  $U \in \Lambda$  we have

$$q|D = \rho q_j = q|D' = q|DU = \rho q_j|U,$$

then  $q_j|U = q_j$ , which means that  $U$  belongs to the group of units

$$E_j = E(q_j) = R(q_j, q_j) \cap \Lambda \quad (2.23)$$

of the quadratic form  $q_j$ . Noticing that the condition  $q|D = \rho q_j$  for a matrix  $D \in K_m^m$  means exactly that  $D \in R(q, \rho q_j)$ , we finally can write the formula (2.12) for each prime  $\rho \in P$  in the form

$$\sum_{j=1}^h \sum_{D \in R^*(q, \rho q_j)/E_j, \substack{D|N \\ D|N}} 1 = c_\rho(q)(1 + \delta(N/\rho)\varepsilon_\rho(q)n(\rho)^{k-1}), \quad (2.24)$$

where

$$R^*(q, \rho q') = R(q, \rho q') \cap \Lambda d_k(\rho)\Lambda. \quad (2.25)$$

**Remark 2.2.** For many Dedekind rings, for example, for rings of integers of finite extensions of the field of rational numbers or fields of  $p$ -adic numbers, there are only finitely many classes of quadratic forms in a given number of variables with a given non-zero determinant (see, for example, [2], Ch. 8–10). For such a ring, the condition (2.20) is automatically fulfilled.

**Remark 2.3.** If  $K$  is a principal ideal domain, it easily follows from the theory of elementary divisors for matrices over  $K$  that

$$R^*(q, \rho q') = R(q, \rho q')$$

for every two quadratic forms  $q$  and  $q'$  in  $m = 2k$  variables over  $K$  with the same determinant  $d \neq 0$  and for each prime element  $\rho$  not dividing  $d$ .

### 3. Rings of automorphs and their action on representations

In order to understand the multiplicative meaning of the lifting formulas (2.24), we introduce here the ring of automorphs of the system of representatives (2.21) and define its action on representations of elements of  $K$  by forms of the system.

First, for an arbitrary set  $S$  and a commutative ring  $A$ , we let  $L_A(S)$  be the  $A$ -linearization of  $S$ , i.e. the free  $A$ -module consisting of all formal finite linear combinations

$$t = \sum_{\alpha} a_{\alpha} \{s_{\alpha}\} \quad a_{\alpha} \in A, \quad s_{\alpha} \in S$$

with coefficients in  $A$  of the symbols  $\{s\}$  corresponding in a one-one way to elements of  $S$ . Further, if we have a pairing of two sets  $S$  and  $X$  into a set  $Z$ , i.e. a mapping

$$(S, X) \rightarrow S \cdot X \subset Z,$$

then by setting

$$\left( \sum_{\alpha} a_{\alpha} \{s_{\alpha}\} \right) \cdot \left( \sum_{\beta} b_{\beta} \{x_{\beta}\} \right) = \sum_{\alpha, \beta} a_{\alpha} b_{\beta} \{s_{\alpha} \cdot x_{\beta}\}, \quad (3.1)$$

$$(L_A(S), L_A(X)) \rightarrow L_A(S) \cdot L_A(X) \subset L_A(Z) \quad (3.2)$$

of their linearizations into  $L_A(Z)$ . The mapping

$$L_A(S) \ni t = \sum_{\alpha} a_{\alpha} \{s_{\alpha}\} \rightarrow c(t) = \sum_{\alpha} a_{\alpha} \in A \quad (3.3)$$

defines clearly a homomorphism of  $A$ -modules

$$c: L_A(S) \rightarrow A$$

which will be called *the coefficient mapping*. The coefficient mappings are compatible with the pairing (3.1), since, by the definition,

$$\begin{aligned} c\left(\left(\sum_{\alpha} a_{\alpha} \{s_{\alpha}\}\right) \cdot \left(\sum_{\beta} b_{\beta} \{x_{\beta}\}\right)\right) &= c\left(\sum_{\alpha, \beta} a_{\alpha} b_{\beta} \{s_{\alpha} \cdot x_{\beta}\}\right) \\ &= \sum_{\alpha, \beta} a_{\alpha} b_{\beta} = \left(\sum_{\alpha} a_{\alpha}\right) \left(\sum_{\beta} b_{\beta}\right) = c\left(\sum_{\alpha} a_{\alpha} \{s_{\alpha}\}\right) \cdot c\left(\sum_{\beta} b_{\beta} \{x_{\beta}\}\right). \end{aligned} \quad (3.4)$$

In what follows, the ring  $A$  will be of no importance. So we shall take  $A$  to be the field  $\mathbb{C}$  of complex numbers.

Coming back to a system of representatives of the form (2.21), we consider the  $\mathbb{C}$ -linearizations

$$L_{ij} = L_{\mathbb{C}}(A_{ij}) \quad (i, j = 1, \dots, h)$$

of the sets

$$A_{ij} = \bigcup_{\mu \in K, \mu \neq 0} R(q_i, \mu q_j)$$

of automorphs of the form  $q_i$  to  $q_j$  with non-zero multipliers  $\mu$ , and  $\mathbb{C}$ -linearizations

$$L_j = L_{\mathbb{C}}(R_j) \quad (j = 1, \dots, h)$$

of the sets

$$R_j = \bigcup_{a \in K} R(q_j, a) = K^m$$

of representations of elements of  $K$  by the form  $q_j$ . It follows from obvious inclusions

$$R(q_i, \mu q_j) \cdot R(q_j, \nu q_r) \subset R(q_i, \mu \nu q_r)$$

and

$$R(q_i, \mu q_j) \cdot R(q_j, a) \subset R(q_i, \mu a)$$

that

$$A_{ij} A_{jr} \subset A_{ir} \quad \text{and} \quad A_{ij} \cdot R_j \subset R_i.$$

Then, by (3.1) and (3.2), we get bilinear pairings

$$(L_{ij}, L_{jr}) \rightarrow L_{ij} \cdot L_{jr} \subset L_{ir} \quad (3.5)$$

and

$$(L_{ij}, L_j) \rightarrow L_{ij} \cdot L_j \subset L_i. \quad (3.6)$$

The pairings (3.5) enable us to define the structure of an associative  $\mathbb{C}$ -algebra on the set

$$\mathbb{L} = L(q_1, \dots, q_h) = \{T = (t_{ij}); t_{ij} \in L_{ij}, i, j = 1, \dots, h\} \quad (3.7)$$

of all  $h \times h$ -matrices with entries in  $L_{ij}$  with respect to the standard matrix operations. Whereas the pairings (3.6) allow us to define a natural linear representation

$$(\mathbb{L}, \mathbb{R}) \rightarrow \mathbb{L} \cdot \mathbb{R} \subset \mathbb{R} \quad (3.8)$$

of the algebra  $\mathbb{L}$  on the space

$$\mathbb{R} = R(q_1, \dots, q_h) = \left\{ X = \begin{pmatrix} x_1 \\ \vdots \\ x_h \end{pmatrix}; x_j \in L_j, j = 1, \dots, h \right\} \quad (3.9)$$

of all  $h$ -columns with entries in  $L_j$ , given by standard multiplication of a matrix by columns. The algebra (3.7) and the space (3.9) will be called respectively the *automorph ring* and the *representation space* of the system  $(q_1, \dots, q_h)$ .

It follows from (3.4) that the coefficient mappings (3.3) define a homomorphism of  $\mathbb{C}$ -algebras

$$c: \mathbb{L} = L(q_1, \dots, q_h) \rightarrow \mathbb{C}_h^h \quad (3.10)$$

and a  $\mathbb{C}$ -linear mapping

$$c: \mathbb{R} = R(q_1, \dots, q_h) \rightarrow \mathbb{C}^h, \quad (3.11)$$

which are compatible with the representation (3.8) of  $\mathbb{L}$  on  $\mathbb{R}$  in the sense that

$$c(T \cdot X) = c(T) \cdot c(X) \quad (T \in \mathbb{L}, X \in \mathbb{R}). \quad (3.12)$$

#### 4. Finite quadratic forms and their multiplicative properties

Here we shall apply the lifting formula (2.24) to study multiplicative properties of an important class of quadratic forms, the finite quadratic forms. A quadratic form  $q$  over a commutative ring  $A$  is said to be *finite* if the set of representations

$$R(q, a) = R_A(q, a) = \{N \in A^m; q(N) = a\},$$

where  $m$  is the number of variables of  $q$ , is finite for each  $a \in K$ . An example of finite forms is given by positive definite forms over rings of integers of totally real finite extensions of the field of rational numbers.

and  $q'$  is a quadratic form satisfying

$$q|D = aq' \quad (4.1)$$

for a non-singular matrix  $D \in A_m^m$  and  $a \in A$ , then the form  $q'$  is also finite.

*Proof.* If a column  $N \in A^m$  satisfies  $q'|N = b$ , then it follows from (4.1) that  $q|DN = ab$ , that is

$$DR(q', b) \subset R(q, ab).$$

Since the last set is finite and the mapping  $N \rightarrow DN$  is an imbedding of  $A^m$  into itself, it follows that the set  $R(q', b)$  is finite.  $\square$

**Lemma 4.2.** *Let  $q$  be a finite quadratic form over a ring  $A$ . Then the set of representations  $R_A(q, q')$  is finite for each quadratic form  $q'$  over  $A$ .*

*Proof.* Let  $q'_{11}, \dots, q'_{nn}$  be the diagonal coefficients of the form  $q'$ . If a matrix  $V$  with columns  $V_1, \dots, V_n$  belongs to  $R(q, q')$ , it follows from  $q|M = q'$  that

$$q|V_1 = q'_{11}, \dots, q|V_n = q'_{nn},$$

which implies that the columns  $V_1, \dots, V_n$  belong to finite sets  $R(q, q'_{11}), \dots, R(q, q'_{nn})$  respectively.  $\square$

Let us return to a system of representatives of the form (2.21), assuming now that  $q$  is a finite form over a Dedekind ring  $K$ . Since the form  $q$  is finite, it follows from (2.19) and Lemma 4.1 that each of the representatives  $q_i$  in (2.21) is finite too. Then, from Lemma 4.2 we conclude that each of the sets  $R(q_i, \mu q_j)$  with  $\mu \in K$  is finite. In particular, the sets  $R^*(q_i, \rho q_j) \subset R(q_i, \rho q_j)$  given by (2.25) and the groups of units (2.23) are finite. This allows us to define elements of the automorph ring (3.7) of the form

$$T^*(\rho) = (T_{ij}^*(\rho)) \in L(q_1, \dots, q_h) \text{ with } T_{ij}^*(\rho) = e_i^{-1} \sum_{D \in R^*(q_i, \rho q_j)} \{D\} \in L_{ij}, \quad (4.2)$$

where

$$e_i = |E_i|$$

is the number of units of  $q_i$ . As to the space of representations (3.9), we introduce the elements of the form

$$R(a) = \begin{pmatrix} R_1(a) \\ \vdots \\ R_h(a) \end{pmatrix} \in R(q_1, \dots, q_h) \text{ with } R_j(a) = e_j^{-1} \sum_{N \in R(q_j, a)} \{N\}. \quad (4.3)$$

The elements we have just introduced have a clear arithmetical meaning since their entries are just *averaged sums* of all representations belonging to corresponding sets.

Now we are in a position to rewrite the formula (2.24) for a finite form  $q$  in terms of the actions of automorph rings on representation spaces. Since clearly  $S(q_i) \subset S(q)$  for each  $i = 1, \dots, h$ , the formula (2.24) is true with  $q_i$  in place of  $q$ . Let us multiply

two sides of the formula with  $q = q_i$  by the symbol  $e_i^{-1}\{N\}$  and sum up over all of  $R(q_i, \rho a)$  for a given  $a \in K$ . For each  $i = 1, \dots, h$  and each  $a \in K$ , we get the relation

$$\begin{aligned} & \sum_{j=1}^h \sum_{N \in R(q_i, \rho a)} e_i^{-1}\{N\} \sum_{\substack{D \in R^*(q_i, \rho q_j) \\ D|N}} 1 \\ &= c_\rho(q) \left( \sum_{N \in R(q_i, \rho a)} e_i^{-1}\{N\} + \varepsilon_\rho(q) n(\rho)^{k-1} \sum_{N \in R(q_i, \rho a) \cap \rho K^m} e_i^{-1}\{N\} \right) \end{aligned} \quad (4.4)$$

(Note that  $c_\rho(q_i) = c_\rho(q)$  and  $\varepsilon_\rho(q_i) = \varepsilon_\rho(q)$  for each  $i = 1, \dots, h$ ). Now, if  $q_i|N = \rho a$ ,  $N = DN'$  with  $D \in R(q_i, \rho q_j)$  and  $N' \in K^m$ , then it follows that  $q_i|N = q_i|DN' = \rho q_j|N'$  from which  $N' \in R(q_j, a)$ . Conversely, if  $N' \in R(q_j, a)$  and  $D \in R(q_i, \rho q_j)$ , it follows that  $N = DN'$  belongs to  $R(q_i, \rho a)$ . Therefore the left side of (4.4) can be written in the form

$$\begin{aligned} & \sum_{j=1}^h \sum_{N \in R(q_i, \rho a)} e_i^{-1}\{N\} \sum_{\substack{D \in R^*(q_i, \rho q_j) \\ N' \in R(q_j, a), DN' = N}} 1 \\ &= \sum_{j=1}^h \sum_{\substack{D \in R^*(q_i, \rho q_j) \\ N' \in R(q_j, a)}} e_i^{-1}\{DN'\} \\ &= \sum_{j=1}^h e_i^{-1} e_j^{-1} \sum_{\substack{D \in R^*(q_i, \rho q_j) \\ N' \in R(q_j, a)}} \{DN'\}, \end{aligned}$$

because, clearly,  $UR(q_j, a) = R(q_j, a)$  for each  $U \in E_j$ . Using the notation (4.2) and (4.3) and the multiplication (3.1), the last expression can be written in the form

$$\begin{aligned} &= \sum_{j=1}^h e_i^{-1} \sum_{D \in R^*(q_i, \rho q_j)} \{D\} \cdot e_j^{-1} \sum_{N' \in R(q_j, a)} \{N'\} \\ &= \sum_{j=1}^h T_{ij}^*(\rho) R_j(a). \end{aligned}$$

The right side of (4.4) in the same notation is clearly equal to

$$\begin{aligned} & c_\rho(q) \left( R_i(\rho a) + \varepsilon_\rho(q) n(\rho)^{k-1} \sum_{\rho N' \in R(q_i, \rho a)} e_i^{-1}\{\rho N'\} \right) \\ &= c_\rho(q) \left( R_i(\rho a) + \varepsilon_\rho(q) n(\rho)^{k-1} \{\rho 1_m\} R_i(a/\rho) \right), \end{aligned}$$

where  $R_i(a/\rho) = 0$  if  $\rho$  does not divide  $a$ . The above considerations show that relation (4.4) can be written in the form

$$\sum_{j=1}^h T_{ij}^*(\rho) R_j(a) = c_\rho(q) (R_i(\rho a) + \varepsilon_\rho(q) n(\rho)^{k-1} \{\rho 1_m\} R_i(a/\rho))$$

valid for  $i = 1, \dots, h$  and for each  $a \in K$ . With the matrix notation (4.2) and (4.3) the relations give us a single matrix relation for each  $a \in K$  and each prime  $\rho$  of  $P$ :

$$[b] = \text{diag}(\{b1_m\}, \dots, \{b1_m\}) \in L(q_1, \dots, q_h), \quad (b \in K), \quad (4.8)$$

and

$$R(a/\rho) = 0 \text{ if } a/\rho \notin K. \quad (4.9)$$

The formula (4.7) is the main goal of our consideration. Let us now join together for convenience of references all of the assumptions made to prove the formula.

**Theorem 4.3.** *Let  $q(X)$  be a quadratic form in an even number  $m = 2k$  of variables with a non-zero determinant  $d$  over a Dedekind ring  $K$ . Suppose that  $q$  satisfies the finiteness of class number condition (2.20) and let  $q_1, \dots, q_h$  be a system of representatives (2.21). Then the formula (4.7) is true for each prime element  $\rho \in K$  of finite norm  $n(\rho)$  which does not divide  $d$  and satisfies the condition (2.14) and each  $a \in K$ , where  $T^*(\rho) \in L(q_1, \dots, q_h)$  is the matrix (4.2),  $R \in R(q_1, \dots, q_h)$  are the columns (4.3) and*

$$c_\rho(q) \neq 0 \text{ and } \varepsilon_\rho(q) = \pm 1$$

are as defined in Theorem 2.1.

**COROLLARY 4.4.**

*With the notation and assumptions of Theorem 4.3, each formal power series of the form*

$$\varphi(t) = \sum_{n=0}^{\infty} R(a\rho^n)t^n$$

*with  $a$  of  $K$  not divisible by  $\rho$  can be formally summed in the form.*

$$\varphi(t) = ([1] - c_\rho(q)^{-1} T^*(\rho)t + \varepsilon_\rho(q)n(\rho)^{k-1}[\rho]t^2)^{-1} \cdot R(a), \quad (4.10)$$

where  $[1] = [1_K]$  and  $[\rho]$  are the elements of the form (4.8) and the inverse on the right is understood in the ring of formal power series in one variable over the ring  $L(q_1, \dots, q_h)$ .

*Proof.* Multiplying the series  $\varphi(t)$  by the matrix  $c_\rho(q)^{-1} T^*(\rho)$  coefficient-wise from the left and using the formula (4.7) with  $a\rho^n$  in place of  $a$ , we obtain

$$\begin{aligned} & c_\rho(q)^{-1} T^*(\rho)\varphi(t) \\ &= \sum_{n=0}^{\infty} (R(a\rho^{n+1}) + \varepsilon_\rho(q)n(\rho)^{k-1}[\rho]R(a\rho^{n-1}))t^n \\ &= (\varphi(t) - R(a))t^{-1} + \varepsilon_\rho(q)n(\rho)^{k-1}[\rho]\varphi(t)t, \end{aligned}$$

since  $R(a\rho^{-1}) = 0$  by (4.9). It follows that

$$\begin{aligned} R(a) &= \varphi(t) - c_\rho(q)^{-1} T^*(\rho)t\varphi(t) + \varepsilon_\rho(q)n(\rho)^{k-1}[\rho]t^2\varphi(t) \\ &= ([1] - c_\rho(q)^{-1} T^*(\rho)t + \varepsilon_\rho(q)n(\rho)^{k-1}[\rho]t^2) \cdot \varphi(t). \end{aligned} \quad (4.11)$$

Since  $[1] = [1_K]$  is the identity element of the ring  $\mathbb{L} = L(q_1, \dots, q_h)$ , it follows that the quadratic polynomial in parentheses on the right is invertible in the ring of

$$\begin{aligned}
& ([1] - c_\rho(q)^{-1} T^*(\rho)t + \varepsilon_\rho(q)n(\rho)^{k-1}[\rho]t^2)^{-1} \\
& = [1] + \sum_{n=1}^{\infty} (c_\rho^{-1}(q) T^*(\rho)t - \varepsilon_\rho(q)n(\rho)^{k-1}[\rho]t^2)^n.
\end{aligned} \tag{4.12}$$

Therefore the relation (4.11) can be written in the form (4.10).  $\square$

Let us now join together the summation formulas (4.10) for all primes  $\rho$  of a countable (or finite) subset  $P'$  of  $P$ :

$$P' = (\rho_1, \rho_2, \dots) \subset P(q).$$

To this end, we let  $S(P')$  be the multiplicative semigroup generated by the unit element  $1 = 1_K$  of  $K$  and all prime elements of  $P'$ . To each element  $b = \rho_1^{n_1} \dots \rho_r^{n_r}$  of  $S(P')$  we associate the monomial

$$t(b) = t(\rho_1^{n_1} \dots \rho_r^{n_r}) = t_1^{n_1} \dots t_r^{n_r}$$

where  $t_1, t_2, \dots$  are commuting independent variables.

#### COROLLARY 4.5.

*With the above notation and assumptions, for each  $a$  of  $K$  which is not divisible by primes  $\rho_1, \rho_2, \dots$ , the following identity for formal power series in  $t_1, t_2, \dots$  with coefficients in the space  $R(q_1, \dots, q_h)$  is valid:*

$$\sum_{b \in S(P')} R(ab)t(b) = \left\{ \prod_{i \geq 1} ([1] - c_i^{-1} T^*(\rho_i)t_i + \varepsilon_i n_i^{k-1} [\rho_i]t_i^2)^{-1} \right\} \cdot R(a) \tag{4.13}$$

where

$$c_i = c_{\rho_i}(q), \quad \varepsilon_i = \varepsilon_{\rho_i}(q) \text{ and } n_i = n(\rho_i).$$

*Proof.* For finite  $P' = P_r = (\rho_1, \dots, \rho_r)$  the relation follows from (4.10) by induction on  $r$ .

If  $P'$  is infinite, we write the relation for finite subsets of the form  $P_r$  and then take the formal coefficient-wise limit as  $r \rightarrow \infty$ .  $\square$

**Remark 4.6.** According to (4.12), each of the  $\rho_i$ -factors on the right can be written as a formal power series in  $t_i$ , say,

$$\begin{aligned}
& ([1] - c_i^{-1} T^*(\rho_i)t_i + \varepsilon_i n_i^{k-1} [\rho_i]t_i^2)^{-1} \\
& = \sum_{n \geq 0} T_i(\rho_i^n) t_i^n,
\end{aligned}$$

where the coefficients

$$T_i(\rho_i^0) = [1], \quad T_i(\rho_i) = c_i^{-1} T^*(\rho_i), \quad T_i(\rho_i^2), \dots$$



$$\sum_{b \in S(P')} T(b) t(b)$$

with

$$T(\rho_1^{n_1} \dots \rho_r^{n_r}) = T_1(\rho_1^{n_1}) \dots T_r(\rho_r^{n_r}) \in L(q_1, \dots, q_h).$$

Then the identity (4.13) means just that

$$R(ab) = T(b) \cdot R(a) \quad (4.14)$$

for each  $a \in K$  not divisible by primes in  $P'$  and each  $b \in S(P')$ . This actually gives us an explicit expression for every one of the averaged sums  $R_1(ab), \dots, R_h(ab)$  of representations of  $ab$  by the forms  $q_1, \dots, q_h$  in terms of  $R_1(a), \dots, R_h(a)$  and the averaged sums  $T_{ij}^*(\rho)$  of automorphs of  $R^*(q_i, \rho q_j)$  for  $i, j = 1, \dots, h$ , where  $\rho$  runs over the prime divisors of  $b$ .

## 5. Numbers of representations by finite forms

The multiplicative properties of representations by finite quadratic forms obtained above imply similar properties of their numbers. For a finite quadratic form  $q$  and an arbitrary quadratic form  $q'$  over a ring  $K$  we shall denote by

$$r(q, q') = |R(q, q')|$$

the number of representations of  $q'$  by  $q$  over  $K$ . According to Lemma 4.2, each of the numbers is finite.

**Theorem 5.1.** *With the notation and assumptions of Theorem 4.3, for each  $\rho \in P$  and each  $a \in K$  the following formula holds:*

$$t^*(\rho) r(a) = c_\rho(q) (r(\rho a) + \varepsilon_\rho(q) n(\rho)^{k-1} r(a/\rho)), \quad (5.1)$$

where

$$t^*(\rho) = (e_i^{-1} r^*(q_i, \rho q_j)) \in \mathbb{C}_h^h \text{ with } r^*(q_i, \rho q_j) = |R^*(q_i, \rho q_j)|$$

and

$$r(a) = \begin{pmatrix} e_1^{-1} r(q_1, a) \\ \vdots \\ e_h^{-1} r(q_h, a) \end{pmatrix} \in \mathbb{C}^h.$$

*Proof.* Applying the coefficient mapping (3.3) to both sides of (4.6) and using (3.4), we get the relations

$$\begin{aligned} \sum_{j=1}^h e_i^{-1} r^*(q_i, \rho q_j) e_j^{-1} r(q_j, a) \\ = c_\rho(q) (e_i^{-1} r(q_i, \rho a) + \varepsilon_\rho(q) n(\rho)^{k-1} e_i^{-1} r(q_i, a/\rho)) \end{aligned} \quad (5.2)$$

for  $i = 1, \dots, h$ . The relations imply (5.1).  $\square$

In the same way as we have proved the corollaries 4.4 and 4.5 or directly from the corollaries, using the mappings (3.11) and (3.10) coefficient-wise and the relations (3.12), we obtain the summation formula

$$\sum_{n \geq 0} r(a\rho^n)t^n = (1_h - c_\rho(q)^{-1}t^*(\rho)t + \varepsilon_\rho(q)n(\rho)^{k-1}t^2)^{-1}r(a)$$

for every  $a \in K$  and  $\rho$  of  $P$  not dividing  $a$ , and the decomposition

$$\sum_{b \in S(P')} r(ab)t(b) = \left\{ \prod_{i \geq 1} \left( 1_h - c_i^{-1}t^*(\rho_i)t_i + \varepsilon_i n_i^{k-1}t_i^2 \right)^{-1} \right\} r(a), \quad (5.3)$$

where notation and assumptions are the same as in (4.10) and (4.13). The last relation can be written in a form similar to (4.14), which shows that it actually gives us an explicit expression of numbers  $r(q_1, ab), \dots, r(q_h, ab)$  of representations of  $ab$  by the forms  $q_1, \dots, q_h$  in terms of  $r(q_1, a), \dots, r(q_h, a)$  and the numbers  $r^*(q_i, \rho q_j)$  for  $i, j = 1, \dots, h$  and  $\rho$  of  $P'$  dividing  $b$ . In the case of quadratic forms over the ring  $\mathbb{Z}$  of rational integers, similar relations could be formerly obtained, only using the theory of Hecke operators on the spaces of theta-series of the quadratic forms (see [3] and [4]).

Let us consider now the following *mean numbers of representations*

$$\bar{r}(a) = \sum_{i=1}^h e_i^{-1} r(q_i, a) r(q_i, 0) \quad (5.4)$$

of elements  $a \in K$  by the forms of a system of representatives (2.21).

**Theorem 5.2.** *Let quadratic forms  $q$  and  $q_1, \dots, q_h$  be the same as in Theorem 4.3, and let  $P = P(q)$  be the set (2.15). Denote by  $S(P)$  the multiplicative semigroup generated by the unit element of the ground ring and by all prime elements of  $P$  and let*

$$\varepsilon: S(P) \rightarrow \{\pm 1\} \text{ and } n: S(P) \rightarrow \mathbb{Z}$$

*be the multiplicative extensions of the mappings*

$$\rho \rightarrow \varepsilon_\rho(q) \text{ and } \rho \rightarrow n(\rho) \quad (\rho \in P)$$

*respectively. Then for each  $b$  of  $S(P)$  and each  $a$  of  $K$  not divisible by prime factors of  $b$  the mean numbers of representations (5.4) satisfy the following relation*

$$\bar{r}(ab) = \left( \sum_{\delta \in S(P), \delta|b} \varepsilon(\delta) n(\delta)^{k-1} \right) \bar{r}(a). \quad (5.5)$$

*Proof.* It is an easy consequence of definitions that the mapping  $D \rightarrow \rho D^{-1}$  for a prime  $\rho$  of  $P$  defines a bijection of each set  $R^*(q_i, \rho q_j)$  onto the set  $R^*(q_j, \rho q_i)$ . In particular, we have the relations

$$r^*(q_i, \rho q_j) = r^*(q_j, \rho q_i) \quad (i, j = 1, \dots, h). \quad (5.6)$$

$$\begin{aligned} & \sum_{j=1}^h e_j^{-1} r^*(q_i, \rho q_j) e_j^{-1} r(q_j, a) r(q_i, 0) \\ &= c_\rho(q)(1 + \varepsilon_\rho(q)n(\rho)^{k-1})r(q_i, 0) \end{aligned} \quad (5.7)$$

valid for  $i = 1, \dots, h$ . Let us multiply now both sides of (5.2) by  $r(q_i, 0)$  and sum up over  $i = 1, \dots, h$ . We obtain the relation

$$\begin{aligned} & \sum_{i,j=1}^h e_i^{-1} r^*(q_i, \rho q_j) e_j^{-1} r(q_j, a) r(q_i, 0) \\ &= c_\rho(q)(\bar{r}(\rho a) + \varepsilon_\rho(q)n(\rho)^{k-1}\bar{r}(a/\rho)). \end{aligned} \quad (5.8)$$

Using (5.7), we can write the left side of (5.8) in the form

$$\begin{aligned} & \sum_{j=1}^h e_j^{-1} r(q_j, a) \sum_{i=1}^h e_i^{-1} r^*(q_i, \rho q_j) r(q_i, 0) \\ &= c_\rho(q)(1 + \varepsilon_\rho(q)n(\rho)^{k-1})\bar{r}(a). \end{aligned}$$

By comparing the expressions and dividing both sides by  $c_\rho(q)$  (note that  $c_\rho(q) \neq 0$  for  $\rho \in P$ ), we get the relation

$$(1 + \varepsilon_\rho(q)n(\rho)^{k-1})\bar{r}(a) = \bar{r}(\rho a) + \varepsilon_\rho(q)n(\rho)^{k-1}\bar{r}(a/\rho).$$

Quite analogously to in the proof of Corollary 4.4, it follows from the last relation that for each  $a \in K$  not divisible by  $\rho$ , the following summation formula

$$\sum_{n \geq 0} \bar{r}(a\rho^n)t^n = (1 - (1 + \varepsilon_\rho(q)n(\rho)^{k-1})t + \varepsilon_\rho(q)n(\rho)^{k-1}t^2)^{-1}\bar{r}(a)$$

holds in the ring of formal power series in one variable over  $\mathbb{C}$ . Since the right side of the formula can be written in the form

$$\begin{aligned} & (1 - t)^{-1}(1 - \varepsilon_\rho(q)n(\rho)^{k-1}t)^{-1}\bar{r}(a) \\ &= \sum_{j \geq 0} \left( \sum_{i=0}^j \varepsilon_\rho(q)^i n(\rho)^{i(k-1)} \right) t^j \bar{r}(a), \end{aligned}$$

the formula implies the relation

$$\bar{r}(a\rho^j) = \left( \sum_{i=0}^j \varepsilon_\rho(q)^i n(\rho)^{i(k-1)} \right) \bar{r}(a)$$

for every  $a$  not divisible by  $\rho$  and  $j = 0, 1, 2, \dots$ . This proves the formula (5.5), if  $b$  is of the form  $b = \rho^j$ . The general case easily follows by induction on the number of different prime divisors of  $b$ .  $\square$

Finally let us have a look at the classical case of positive definite quadratic forms over the ring  $\mathbb{Z}$  with determinant 1. In this case, there are only finitely many classes

of quadratic forms  $q$  in a given number  $m$  of variables and we can take  $q_1, \dots, q_h$  to be a system of representatives for the classes. Since in this case  $m = 2k \equiv 0 \pmod{8}$  (see, for example [5], Ch. 5), it implies that

$$\varepsilon_p(q) = \varepsilon_p(q_i) = 1$$

for each prime  $p$ , and so  $P$  is the set of all prime numbers and  $S(P)$  is the multiplicative semigroup  $\mathbb{N}$  of all natural numbers. The mean number (5.4) of representations of a natural number  $b$  by the forms  $q_1, \dots, q_h$  turns into

$$\bar{r}(b) = \sum_{i=1}^h e_i^{-1} r(q_i, b),$$

and the formula (5.5) for  $a = 1$  turns into the relation

$$\bar{r}(b) = \left( \sum_{\delta \in \mathbb{N}, \delta|b} \delta^{k-1} \right) \bar{r}(1). \quad (5.9)$$

To get a complete picture, it would be nice to find also a formula for  $\bar{r}(1)$ . Unfortunately we can do it only using transcendental means. We shall give here a brief account of computations; for details see, for example, [5], Ch. 7 or [6], Ch. 4, 6. The theta-series

$$\Theta_j(z) = \sum_{n \geq 0} r(q_j, n) \exp(2\pi i n z) \quad (z = x + iy, y > 0)$$

of the forms  $q_1, \dots, q_h$  belong to the space  $\mathcal{M}_k$  of holomorphic modular forms of weight  $k = m/2$  relative to the modular group  $SL_2(\mathbb{Z})$ , and so does their linear combination

$$\bar{\Theta}(z) = \sum_{j=1}^h e_j^{-1} \Theta_j(z) = \sum_{n \geq 0} \bar{r}(n) \exp(2\pi i n z),$$

which according to (5.9) can be written in the form

$$\bar{\Theta}(z) = \bar{r}(0) + \bar{r}(1) \sum_{n \geq 1} \left( \sum_{\delta|n} \delta^{k-1} \right) \exp(2\pi i n z).$$

On the other hand, the space  $\mathcal{M}_k$  contains also the so-called Eisenstein series  $E_k(z)$  with the Fourier expansion of the form

$$E_k(z) = 1 + \gamma_k \sum_{n \geq 1} \left( \sum_{\delta|n} \delta^{k-1} \right) \exp(2\pi i n z)$$

where

$$\gamma_k = \frac{(2\pi)^k}{(k-1)! \zeta(k)} \quad \text{with} \quad \zeta(k) = \sum_{n \geq 1} n^{-k},$$

and so the space contains also the linear combination

$$\gamma_k \bar{\Theta}(z) - \bar{r}(1) E_k(z) = \gamma_k \bar{r}(0) - \bar{r}(1),$$

which is a constant. But the space  $\mathcal{M}_k$  contains no constants except zero and so

$$\bar{r}(1) = \gamma_k \bar{r}(0) = \frac{(2\pi)^k}{(k-1)!\zeta(k)} \sum_{j=1}^h e_j^{-1}.$$

Substituting it in (5.9), we finally obtain the formula

$$\left( \sum_{j=1}^h e_j^{-1} \right)^{-1} \sum_{j=1}^h e_j^{-1} r(q_j, b) = \frac{(2\pi)^k}{(k-1)!\zeta(k)} \sum_{\delta|b} \delta^{k-1}$$

valid for all natural numbers  $b$ , which is a finite form of the famous Siegel theorem [7] on mean numbers of representations for positive quadratic forms over  $\mathbb{Z}$  with determinant 1.

## 6. Concluding remarks

The case when the sets of representations  $R(q, q')$  are, in general, infinite can be considered in a similar way, provided that the sets  $E(q) \setminus R(q, q')$  of cosets modulo the corresponding group of units  $E(q) = R(q, q) \cap \Lambda$  are finite, which often happens. In this case, instead of linear combinations of representations  $D$  of  $R(q, q')$ , one should consider corresponding linear combinations of their cosets  $E(q)D$  modulo the group of units. For more, see [1], § 5.

The technique of lifting of solutions of quadratic congruences modulo prime elements, unfortunately, does not work in the case of quadratic forms in an odd number of variables. Recently F A Andrianov (Jr.) has succeeded in considering multiplicative properties of positive definite quadratic forms in an odd number of variables over the ring  $\mathbb{Z}$ , using a technique of lifting of solutions of quadratic congruences modulo squares of prime numbers. A generalization to finite forms over Dedekind rings goes through without any problem. The corresponding papers by F A Andrianov are due to appear in *Zapiski Nauchn. Sem. St. Petersburg Otdel. Steklov Math. Inst.*, vol. 212 1993/94.

## Acknowledgement

This research was carried out and written in the spring of 1993 during the author's stay at Sonderforschungsbereich 170 "Geometrie und Analysis" at Mathematical Institute of Göttingen University. The author would like to take this opportunity to thank the SFB-Board, and especially Professor Ulrich Christian, for their kind invitation, hospitality, and very stimulating research environment.

## References

- [1] Andrianov A N, Automorphic factorizations of solutions of quadratic congruences and their applications, *Algebra and Analysis* 5 (1993), No. 5 (Russ.); Engl. transl.: *St. Petersburg Math. J.* 5 1-46 (1993)
- [2] O'Meara O T, Introduction to quadratic forms (*Grundlehren Math. Wiss.* Band 117, Springer, Berlin-Heidelberg) (1963)

- [3] Andrianov A N, Integral representations of quadratic forms by quadratic forms: multiplicative properties, In *Proc. Int. Cong. Math. Warszawa*, 1 1983, Vol. I, pp. 465–474
- [4] Andrianov A N, Composition of solutions of quadratic Diophantine equations, *Russ. Math. Surv.* **46:2** (1991) 1–44
- [5] Serre J P, *Cours d'Arithmétique*, (Press Universitaire de France, Paris) (1970)
- [6] Ogg A, *Modular forms and Dirichlet series*, (Benjamin, New York – Amsterdam) (1969)
- [7] Siegel C L, Über die analytische Theorie der quadratischen Formen, Teil I, *Ann. Math.* **36** (1935) 527–606

# Non-surjectivity of the Clifford invariant map

R PARIMALA and R SRIDHARAN

School of Mathematics, Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400 005, India

Dedicated to the memory of Professor K G Ramanathan

**Abstract.** The question whether there exists a commutative ring  $A$  for which there is an element in the 2-torsion of the Brauer group not represented by a Clifford algebra was raised by Alex Hahn. Such an example is constructed in this paper and is arrived at using certain results of Parimala–Sridharan and Parimala–Scharlau which are also reviewed here.

**Keywords.** Quadratic forms; Clifford invariant; Brauer group; canonical class.

## 1. Introduction

A celebrated theorem of Merkurjev [M] asserts that if  $k$  is a field, every element in the 2-torsion of the Brauer group of  $k$  is represented by the Clifford algebra of a quadratic form over  $k$ . The following question was raised by Alex Hahn: Does there exist a commutative ring  $A$  and an element in the 2-torsion of the Brauer group of  $A$  which cannot be represented by the Clifford algebra of a quadratic form over  $A$ ? Examples of smooth projective curves  $X$  over local fields for which the Clifford algebra classes do not fill up the 2-torsion in the Brauer group of  $X$  were given in [P–Sr]. These were arrived at in the following manner: While comparing the graded Witt ring of a curve with graded unramified cohomology ring, necessary and sufficient conditions were given in [P–Sr], under which the ‘Clifford invariant’ map surjects on to the 2-torsion in the Brauer group, for a smooth projective curve over a local field. In a joint work with Scharlau [P–S], the above mentioned conditions were shown to be equivalent to the condition that the canonical class of the curve is “even” for smooth projective hyperelliptic curves over local fields. An explicit condition was also given in this case as to when the canonical class is even. This leads to the requisite examples of smooth projective curves over local fields for which the Clifford invariant map is not surjective. In this paper we review these results and use them to construct an affine algebra  $A$  for which the Clifford invariant map is not surjective, thus answering the question of Alex Hahn.

Throughout this paper,  $k$  denotes a field of characteristic not 2.

## 2. The Clifford invariant map

Let  $X$  be a smooth integral variety over field  $k$ . A *quadratic space* on  $X$  is a locally free sheaf  $\mathcal{E}$  on  $X$  together with a self-dual isomorphism  $q: \mathcal{E} \rightarrow \mathcal{E}^\vee = \text{Hom}(\mathcal{E}, \mathcal{O}_X)$ . If

algebras on  $X$  and its class in the Brauer group of  $X$  is called the *Clifford invariant* of  $(\mathcal{E}, q)$ , denoted by  $e_2(q)$  (the *second invariant*, the first being the discriminant). Let  $W(X)$  denote the Witt group of  $X$ , namely the quotient of the Grothendieck group of quadratic spaces over  $X$  under orthogonal sum modulo the subgroup generated by metabolic spaces [K]. Let  $I_2(X)$  denote the subgroup of  $W(X)$  generated by spaces of even rank and trivial discriminant. The Clifford invariant is well defined on Witt equivalence classes [Kn-Oj] and defines a homomorphism  $e_2: I_2(X) \rightarrow {}_2Br(X)$ ,  ${}_2Br(X)$  denoting the 2-torsion subgroup of the Brauer group of  $X$ . If  $X = \text{Spec } k$ , the theorem of Merkurjev mentioned earlier assures that  $e_2$  is surjective. The next non-trivial case is that of a smooth integral curve  $X$ .

We recall from [P-Sr] some results concerning this question for curves over local fields. We look at the case when  $k$  is a non-archimedean local field. Let  $X$  be a smooth integral curve over  $k$  and  $X^{(1)}$  the set of closed points of  $X$ . We have an exact sequence

$$0 \rightarrow W(X) \xrightarrow{i} W(k(X)) \xrightarrow{(\delta_x)} \bigoplus_{x \in X^{(1)}} W(k(x)),$$

where  $i$  is the restriction to the generic point and  $\delta_x: W(k(X)) \rightarrow W(k(x))$  is a residue homomorphism defined with respect to a choice of the parameter for the discrete valuation corresponding to  $x \in X^{(1)}$ . The powers of the ideal  $I(k(X))$  of even rank forms in  $W(k(X))$  induces a filtration

$$I_n(X) = W(X) \cap I^n(k(X))$$

on  $W(X)$ . The above exact sequence respects this filtration and yields the following exact sequence

$$0 \rightarrow I_n(X) \rightarrow I^n(k(X)) \rightarrow \bigoplus_{x \in X^{(1)}} I^{n-1}(k(x)).$$

Since the cohomological dimension of  $k$  is 2, by a theorem of [A-E-J], there exist well defined surjective homomorphisms  $e_n: I^n(k(X)) \rightarrow H^n(k(X))$ , with kernel precisely  $I^{n+1}(k(X))$  (we note that  $e_2$  is simply the Clifford invariant map). The same is true for  $I^n(k(x))$  for  $x \in X^{(1)}$ . We have a homomorphism  $\partial = (\partial_x): H^n(k(X)) \rightarrow \bigoplus_{x \in X^{(1)}} H^{n-1}(k(x))$

whose kernel is the unramified cohomology group  $H^0(X, \mathcal{H}^n)$ , where  $\mathcal{H}^n$  denotes the Zariski sheaf associated with the presheaf  $U \mapsto H^n_{\text{et}}(U, \mu_2)$  [B-O]. The following diagram of exact rows

$$\begin{array}{ccccccc} 0 & \rightarrow & I_n(X) & \rightarrow & I^n(k(X)) & \xrightarrow{(\delta_x)} & \bigoplus_{x \in X^{(1)}} I^{n-1}(k(x)) \\ & & \downarrow e_n & & \downarrow e_n & & \downarrow (e_{n-1}) \\ 0 & \rightarrow & H^0(X, \mathcal{H}^n) & \rightarrow & H^n(k(X)) & \xrightarrow{(\partial_x)} & \bigoplus_{x \in X^{(1)}} H^{n-1}(k(x)) \end{array}$$

is commutative [P<sub>3</sub>]. By [A-E-J], for  $n \geq 4$ ,  $I^n(k(X)) = 0$ , so that  $e_3: I^3(k(X)) \rightarrow H^3(k(x))$  is an isomorphism. Further, for  $x \in X^{(1)}$ ,  $e_2: I^2(k(x)) \rightarrow H^2(k(x))$  is also an isomorphism. This implies that  $e_3: I_3(X) \rightarrow H^0(X, \mathcal{H}^3)$  is an isomorphism. Further,



the cokernel of the map  $I^3(k(X)) \xrightarrow{\delta} \bigoplus_{x \in X^{(1)}} I^2(k(x))$  is isomorphic to the cokernel of  $\partial: H^3(k(X)) \rightarrow \bigoplus_{x \in X^{(1)}} H^2(k(x))$ , which, by Bloch–Ogus theory [B–O], is isomorphic to  $H^1(X, \mathcal{H}^3)$ , which is a subgroup of  $H_{\text{et}}^4(X, \mu_2)$ . If  $X$  is not projective, since  $cd_2 k = 2$ , an analysis of the Hochschild–Serre spectral sequence yields that  $H_{\text{et}}^4(X, \mu_2) = 0$ , so that the map  $\delta: I^3(k(X)) \rightarrow \bigoplus_{x \in X^{(1)}} I^2(k(x))$  is surjective. The following commutative diagram of exact rows and columns

$$\begin{array}{ccccccc}
 & & 0 & & 0 & & 0 \\
 & & \downarrow & & \downarrow & & \downarrow \\
 0 & \rightarrow & I_3(X) & \rightarrow & I^3(k(X)) & \rightarrow & \bigoplus_{x \in X^{(1)}} I^2(k(x)) \rightarrow 0 \\
 & & \downarrow & & \downarrow & & \downarrow \\
 0 & \rightarrow & I_2(X) & \rightarrow & I^2(k(X)) & \rightarrow & \bigoplus_{x \in X^{(1)}} I(k(x)) \\
 & & \downarrow e_2 & & \downarrow e_2 & & \downarrow (e_1) \\
 0 & \rightarrow & H^0(X, \mathcal{H}^2) & \rightarrow & H^2(k(X)) & \rightarrow & \bigoplus_{x \in X^{(1)}} H^1(k(x)) \\
 & & & & \downarrow & & \downarrow \\
 & & & & 0 & & 0
 \end{array}$$

for an affine curve  $X$  over  $k$ , shows that  $e_2: I_2(X) \rightarrow H^0(X, \mathcal{H}^2)$  is surjective. By purity theorem ([Gr], Prop. 2.1),  $H^0(X, \mathcal{H}^2) \simeq {}_2\text{Br}(X)$  and  $e_2$  is the Clifford invariant map. We thus have proved the following

**Theorem 1.** ([P–Sr], Th. 4.4) *If  $X$  is an affine curve over a local field, the Clifford invariant map  $e_2: I_2(X) \rightarrow {}_2\text{Br}(X)$  is surjective.*

Suppose now that  $X$  is a projective curve which has a  $k$ -rational point. Let  $x_0 \in X(k)$ . By ([A]) Satz. 4.16), we have a complex

$$H^i(k(X)) \xrightarrow{\partial} \bigoplus_{x \in X^{(1)}} H^{i-1}(k(x)) \xrightarrow{\text{cores}} H^{i-1}(k).$$

If  $\alpha \in H^i(k(X))$  is such that  $\partial(\alpha)$  has at most one non-zero component at  $x_0$ , then  $\partial(\alpha) = 0$  since  $\text{cores}: H^{i-1}(k(x_0)) \rightarrow H^{i-1}(k)$  is an isomorphism. Thus, if  $Y = X \setminus \{x_0\}$ ,  $H^0(X, \mathcal{H}^i) \simeq H^0(Y, \mathcal{H}^i)$  for all  $i$ . In particular,  ${}_2\text{Br}(X) \simeq {}_2\text{Br}(Y)$ . In the following commutative diagram

$$\begin{array}{ccc}
 I_3(X) & \simeq & H^0(X, \mathcal{H}^3) \\
 \downarrow & & \downarrow \\
 I_3(Y) & \simeq & H^0(Y, \mathcal{H}^3)
 \end{array}$$

where the vertical arrows are induced by the inclusion  $Y \subset X$ , all arrows, except possibly the left vertical arrow, are isomorphisms and hence the restriction map  $I_3(X) \rightarrow I_3(Y)$  is an isomorphism. Finally, we look at the following commutative

$$\begin{array}{ccccccc}
0 & \rightarrow & I_3(X) & \rightarrow & I_2(X) & \xrightarrow{e_2} & {}_2\text{Br}(X) \\
& & \downarrow & & \downarrow & & \downarrow \\
0 & \rightarrow & I_3(Y) & \rightarrow & I_2(Y) & \xrightarrow{e_2} & {}_2\text{Br}(Y) \rightarrow 0.
\end{array}$$

The map  $e_2: I_2(X) \rightarrow {}_2\text{Br}(X)$  is surjective if and only if the map  $I_2(X) \rightarrow I_2(Y)$  is surjective. This leads to the following definition

## DEFINITION

Let  $X$  be a smooth projective curve over a field  $k$ . We say that  $X$  has *extension property* (for quadratic spaces) if there exists a rational point  $x_0 \in X(k)$  such that every quadratic space over  $X \setminus x_0$  extends to  $X$ .

**Theorem 2.** ([P-Sr], Th. 4.4). *Let  $X$  be a smooth projective curve over a local field with a rational point. The Clifford invariant map  $e_2: I_2(X) \rightarrow {}_2\text{Br}(X)$  is surjective if and only if  $X$  has extension property.*

*Proof.* Let  $x_0 \in X(k)$  and  $Y = X \setminus \{x_0\}$ . We need only to verify that if the map  $I_2(X) \rightarrow I_2(Y)$  is surjective then every quadratic space on  $Y$  extends to  $X$ . Suppose  $I_2(X) \rightarrow I_2(Y)$  is surjective. Let  $q$  be any quadratic space on  $Y$ . We check that the second residue  $\delta_{x_0}(q) = 0$ . If rank  $q$  is odd, we replace  $q$  by  $q \perp \langle 1 \rangle$  and assume that rank  $q$  is even. The space disc  $q \in H^1(Y, \mu_2) = H^1(X, \mu_2)$ , so that disc  $q$  is nonsingular on  $X$ . Replacing  $q$  by  $q \perp \langle -1, \text{disc } q \rangle$  which has the same residue as  $q$  at any point, we assume that disc  $q$  is trivial so that  $q \in I_2(Y)$ . Then by assumption,  $q$  extends to  $X$ .  $\square$

**Remark 1.** The extension property for a smooth projective curve  $X$  over any field could be defined as above with respect to a given rational point  $x_0$ . It is interesting to study the equivalence classes on  $X(k)$  defined by  $x \sim y$  if and only if  $X$  has extension property 'with respect to  $x$ ' is equivalent to  $X$  has extension property 'with respect to  $y$ '. The theorem implies that for a smooth projective curve over a local field there is only one equivalence class: i.e., the extension property defined with respect to  $x$  does not depend on the choice of  $x$ .

**Remark 2.** It may be shown that the map  $e_2: I_2(X) \rightarrow {}_2\text{Br}(X)$  is surjective if and only if every  $\xi \in {}_2\text{Br}(X)$  is the class of a Clifford algebra of some even rank quadratic space over  $X$ . In fact if  $\xi \in {}_2\text{Br}(X)$  is such that  $\xi = C(\mathcal{E}, q)$  with rank  $q$  even,  $\xi = e_2(q')$  where  $q' = q \perp \langle -1, \text{disc } q \rangle \in I_2(X)$ .

## 3. Canonical class of a curve and extension property for quadratic spaces

We now go on to analyse when the extension property holds for a smooth projective curve  $X$ . In view of a theorem of Geyer–Harder–Knebusch–Scharlau [G–H–K–S],

a sufficient condition for the extension property to hold for  $X$  is that the canonical line bundle  $\Omega_X$  on  $X$  even; i.e.,  $\Omega_X$  is a square in  $\text{Pic } X$ . This is due to a certain reciprocity for quadratic spaces on  $X$  if  $\Omega_X$  is even. We explain this reciprocity.

Let  $X$  be a smooth projective curve over a field  $k$  with  $\text{char } k \neq 2$  and  $k$  perfect. Let  $\Omega_X = \Omega$  be the canonical sheaf on  $X$ . We can define the Witt group  $W(X, \Omega)$  of quadratic spaces on  $X$  with values in the line bundle  $\Omega$ . There are canonical residue homomorphisms

$$\partial_x: W(k(X), \Omega_{k(X)}) \rightarrow W(k(x))$$

for each closed point  $x$  of  $X$ . Any non-singular quadratic form  $q$  over  $k(X)$  with values in  $\Omega_{k(X)}$  may be written as

$$q = q_1 d\pi \perp q_2 \left( \frac{d\pi}{\pi} \right).$$

with  $q_1$  and  $q_2$  regular over  $\mathcal{O}_{X,x}$  and  $\pi$  a parameter at  $x$ . Then  $\partial_x(q)$  is the reduction of  $q_2$  modulo  $\pi$ , which is independent of the choice of  $\pi$ . It is proved in [G-H-K-S] that the sequence

$$W(k(X), \Omega_{k(X)}) \xrightarrow{(\partial_x)} \bigoplus_{x \in X^{(1)}} W(k(x)) \xrightarrow{s} W(k)$$

is a complex,  $s$  being the transfer induced by the trace map. If  $\Omega_X$  is a square in  $\text{Pic } X$ , we have isomorphisms of the following complexes

$$\begin{array}{ccccccc} 0 & \rightarrow & W(X, \Omega_X) & \rightarrow & W(k(X), \Omega_{k(X)}) & \xrightarrow{(\partial_x)} & \bigoplus_{x \in X^{(1)}} W(k(x)) \xrightarrow{s} W(k) \\ & & z \downarrow & & z \downarrow & & z \downarrow \\ 0 & \rightarrow & W(X) & \rightarrow & W(k(X)) & \xrightarrow{(\partial_x)} & \bigoplus_{x \in X^{(1)}} W(k(x)) \end{array}$$

for  $\delta_x$  defined through certain choice of parameters at  $x, x \in X^{(1)}[P_2]$ . In particular, if a form  $q \in W(k(X))$  has non-zero residue at possibly one rational point  $x_0$ , then the residue at this rational point is necessarily zero. Hence  $q$  on  $X \setminus x_0$  extends to  $X$ . Thus if  $\Omega_X$  is even,  $X$  has extension property.

One is led into analysing when  $\Omega_X$  is even. This is purely a rationality question, since over the algebraic closure of  $k$ , degree  $\Omega_X$  is even;  $\text{Pic}^0 X$  being divisible,  $\Omega_{X_{\bar{k}}}$  is a square. In particular, there is extension property for curves over algebraically closed fields. In [P-S], a necessary and sufficient condition was given as to when  $\Omega_X$  is even for hyperelliptic curves over any field (see also [Su]). It so happens that for smooth projective hyperelliptic curves over *local* fields, extension property is *equivalent* to  $\Omega_X$  being even.

**Theorem 3.** ([P-S], Th. 2.4). *Let  $k$  be a local field with  $\text{char } k \neq 2$  and  $X$  a smooth projective hyperelliptic curve of genus at least two with  $X(k) \neq \emptyset$ . Then the following are equivalent:*

- (1)  $\Omega_X$  is even.
- (2)  $X$  has extension property.

- (3) genus  $X$  is odd or genus  $X$  is even and  $X$  satisfies one of the following for a double covering  $\pi: X \rightarrow \mathbf{P}^1$ ,
- (a)  $\pi$  has a ramification point of odd degree.
  - (b) All ramification points of  $\pi$  have even degree and there is a quadratic extension of  $k$  which is contained in the residue fields of all ramification points of  $\pi$ .

**Remark 3.** The conditions (a) and (b) of (3) are intrinsic for  $X$  since the covering  $\pi: X \rightarrow \mathbf{P}^1$  is unique up to isomorphism, genus of  $X$  being at least 2. Condition (3) (b) as stated in ([P-S]) includes a further condition, namely that for some choice of a rational point  $\infty$  for  $\mathbf{P}^1$ , if  $\infty$  is inert for  $\pi$  and  $k(\infty) = k(\sqrt{\eta})$ , then  $\eta$  is a norm from  $\ell$ . However, in our situation,  $X(k) \neq \emptyset$  and we may choose  $\infty$  to be lying below some rational point of  $X$  so that  $\infty$  is split and the extra condition is vacuous.

Now it is clear as to how to construct examples of curves  $X$  over a local field  $k$  such that  $\Omega_X$  is not even. Let  $k = \mathbb{Q}_3$ ,  $p_1(t)$  an irreducible polynomial of degree 2 and  $p_2(t)$  an irreducible polynomial of degree 4 over  $\mathbb{Q}_3$  such that  $\mathbb{Q}_3[t]/p_1(t)$  is totally ramified and  $\mathbb{Q}_3[t]/(p_2(t))$  is unramified. (e.g.  $p_1(t) = t^2 - 3$ ,  $p_2(t) = t^4 + t^3 + t^2 + t + 1$ ). The hyperelliptic curve  $y^2 = p_1(t)p_2(t)$  of genus 2 has two points of ramification. Clearly the residue fields at these points do not have a common quadratic extension since one is unramified and the other totally ramified. Hence by the above theorem,  $\Omega_X$  is not even. Further there are choices for  $p_1(t)$  and  $p_2(t)$  (e.g., the example above) for which  $X(k) \neq \emptyset$ . One can even construct explicitly, a quadratic space over  $X \setminus \text{a rational point}$ , which does not extend to  $X$ .

#### 4. The example

In this section, we construct a commutative ring  $A$  for which there is an element in  ${}_2\text{Br}(A)$  which is not the class of the Clifford algebra of any quadratic space over  $A$ .

We recall that a *locally trivial affine fibration*  $\pi: Y \rightarrow X$  of schemes is one for which, locally, at each point  $x \in X$ ,  $\pi: Y \times_X \text{Spec } \mathcal{O}_{X,x} \rightarrow \text{Spec } \mathcal{O}_{X,x}$  is isomorphic to the projection  $A' \times_X \text{Spec } \mathcal{O}_{X,x} \rightarrow \text{Spec } \mathcal{O}_{X,x}$ .

**Theorem 4.** *Let  $X$  be a smooth projective curve over a field  $k$  and  $\pi: \text{Spec } A \rightarrow X$  a locally trivial affine fibration. If  $\xi \in {}_2\text{Br}(X)$  is not the Clifford invariant of any quadratic space over  $X$ , then  $\pi^*\xi \in {}_2\text{Br}(A)$  is not the Clifford invariant of any quadratic space over  $A$ .*

*Proof.* Suppose  $q$  is a quadratic space of even rank over  $A$  such that  $\mathcal{C}(q)$  defines the class of  $\pi^*\xi$ . We may assume, by adding a hyperbolic plane, if necessary, that  $q$  contains a hyperbolic plane. We may also assume, by adding  $\langle -1, \text{disc } q \rangle$ , if necessary, that  $[q] \in I_2(A)$  and  $e_2(q) = \pi^*\xi$ . Since the fibration  $\pi: \text{Spec } A \rightarrow X$  is locally trivial, for each closed point  $x \in X^{(1)}$ ,  $\pi: \text{Spec } A \times_X \text{Spec } \mathcal{O}_{X,x} \rightarrow \text{Spec } \mathcal{O}_{X,x}$  is given by  $\mathcal{O}_{X,x} \hookrightarrow \mathcal{O}_{X,x} [T_1, \dots, T_r]$  where  $T_i$  are indeterminates. Since  $\mathcal{O}_{X,x}$  is a discrete valuation ring, in view of ([P<sub>1</sub>], Th. 3.2), there exists a quadratic space  $q_x$  over  $\mathcal{O}_{X,x}$  such that  $\pi^*q_x \simeq q|_{\text{Spec } A \times_X \text{Spec } \mathcal{O}_{X,x}}$ . The spaces  $q_x$  become isometric, at the generic point  $x_0$  of  $X$  to the space  $q_0$  defined by  $\pi^*q_0 \simeq q|_{\text{Spec } A \times_X \text{Spec } k(X)}$ . (Observe that  $\text{Spec } A \times_X \text{Spec } k(X) \simeq \text{Spec } k(X)[T_1, T_2, \dots, T_r]$  and  $q$  restricted to  $\text{Spec } A \times_X \text{Spec } k(X)$  comes from the space  $q_0$  over  $k(X)$   $q$  being isotropic ([Oj]). Through a typical dimension one argument, one sees that there is a quadratic space  $q_1$  over  $X$ , which, when restricted

to the generic point  $x_0$  of  $X$  becomes isometric to  $q_0$ . We note that  $\text{disc } q_1 = 1$ , since, locally,  $\text{disc } q_1 \otimes \mathcal{O}_{X,x} = \text{disc } \pi^*(q_1 \otimes \mathcal{O}_{X,x})$  is trivial, so that  $q_1 \in I_2(X)$ . We show that  $e_2(q_1) = \xi$  which leads to a contradiction through our choice of  $\xi$ . Since  $e_2$  is functorial and the map  ${}_2\text{Br}(X) \rightarrow {}_2\text{Br}(k(X))$  is injective ([Gr], Prop. 2.1], it suffices to show that  $e_2(q_{1(k(x))}) = \xi_{k(X)}$ . The map  $\pi^*: \text{Br}(k(X)) \rightarrow \text{Br}(\text{Spec } A \times_X \text{Spec } k(X))$  is a (split) injection and hence it suffices to show that  $\pi^*(e_2(q_{1(k(x))})) = \pi^*(\xi_{k(X)})$ . We have

$$\begin{aligned}\pi^*(e_2(q_{1(k(x))})) &= \pi^*e_2(q_0) \\ &= e_2(\pi^*q_0) \\ &= e_2(q_{k(X)})\end{aligned}$$

However, by choice,  $e_2(q) = \pi^*\zeta$  so that  $e_2(q_{k(X)}) = \pi^*\zeta_{k(X)}$ . Thus  $\pi^*e_2(q_{1(k(x))}) = \pi^*(\xi_{k(X)})$ , leading to a contradiction.

Let  $X$  be a smooth projective curve over a field  $k$ . We recall the construction of a locally trivial affine fibration  $T: W \rightarrow X$  with  $W$  affine, due to Jouanolou [J]. Let  $j: X \hookrightarrow \mathbb{P}_k^r = \mathbf{P}$  be a closed immersion. Let  $W(r)$  be the Stiefel variety over  $k$  given by the equation  $\{E^2 = E, \text{Trace } E = 1\}$  where  $E$  is the  $(r+1) \times (r+1)$  generic matrix  $(x_{ij})$  over  $k$ . Clearly  $W(r)$  is affine and is a principal homogeneous space for the vector bundle  $\text{Hom}(\mathcal{F}, \mathcal{L})$  where  $\mathcal{L}$  is the canonical line bundle  $\mathcal{O}(-1)$  on  $\mathbf{P}$  and  $\mathcal{F}$  is defined by the exact sequence

$$0 \rightarrow \mathcal{L} \rightarrow \mathcal{O}_{\mathbf{P}}^{r+1} \rightarrow \mathcal{F} \rightarrow 0.$$

The natural map  $\pi: W(r) \rightarrow \mathbf{P}$ , given by  $E \mapsto \text{image}(E)$  is a locally trivial affine fibration. We define  $\pi: W \rightarrow X$  by the Cartesian square

$$\begin{array}{ccc} W & \rightarrow & W(r) \\ \pi \downarrow & & \downarrow \pi \\ X & \xrightarrow{j} & \mathbf{P} \end{array}$$

The map  $\pi: W \rightarrow X$  is a locally trivial affine fibration with each fibre an affine  $r$ -space and  $W$  is affine.

We now give the promised example. Let  $X$  be the smooth projective hyperelliptic curve over  $\mathbb{Q}_3$  defined by  $y^2 = (t^2 - 3)(t^4 + t^3 + t^2 + t + 1)$ . Let  $\pi: \text{Spec } A \rightarrow X$  be an  $A^2$ -fibration described above. Then  $A$  is an affine algebra over  $\mathbb{Q}_3$  of dimension 3. By the above Theorem and the discussion at the end of §2, there is a Brauer class in  $A$  which is not the class of the Clifford algebra of any quadratic space over  $A$ .

*Remark 4.* In view of Theorem 1, for any affine curve over a local field, the Clifford invariant map is surjective. The following example, pointed out to us by Kapil Paranjape, gives a locally trivial  $A^1$ -fibration of a smooth projective curve  $C$  with total space affine. Let  $Y$  be the complement of a non-constant section in  $C \times \mathbb{P}^1$  and  $\pi: Y \rightarrow C$  the projection. This once again leads to examples of affine surfaces over  $p$ -adic fields for which the Clifford invariant map is not surjective. For an arbitrary ground field  $k$ , there are even perhaps examples of affine curves for which the Clifford invariant map is not surjective.

## References

- [A] Arason J Kr, Cohomologische invarianten quadratischer formen, *J. Algebra* **36** (1975) 479–491
- [A–E–J] Arason J Kr, Elman R and Jacob B, Fields of cohomological 2-dimension three, *Math. Ann.* **274** (1986) 649–657
- [B–O] Bloch S and Ogus A, Gersten's conjecture and the homology of schemes, *Ann. Sci. Éc. Norm. Supér* **7** (1974) 181–202i
- [G–H–K–S] Geyer W D, Harder G, Knebusch M and Scharlau W, Ein Residuensatz für symmetrische Bilinear Formen, *Invent Math.* **11** (1970) 319–328
- [Gr] Grothendieck A, Le groupe de Brauer III: Exemples et complements, In *Dix Exposés sur la cohomologie des schémas* 1968 (North-Holland, Amsterdam) pp. 46–188
- [J] Jouanolou J-P, Une suite exacte de Mayer-Vietoris en  $K$ -Théorie algebrique, *SLN* **341**, (1973) 293–316
- [K] Knebusch M, Symmetric bilinear forms over algebraic varieties, Conference on Quadratic Form, *Queen's Papers in Pure and Applied Mathematics*, **46** 1977
- [Kn–Oj] Knus M-A and Ojanguren M, The Clifford algebra of a metabolic space, *Arch. der Math.*, Basel **56** (1991) 440–441
- [L] Lam T-Y, The algebraic theory of quadratic forms (1973) (Benjamin)
- [M] Merkurjev A, On the norm residue symbol of degree 2, *Sov. Math. Dokl.* **24** (1981) 546–551
- [Oj] Ojanguren M, Formes quadratiques sur les algebres de polynômes, *Comptes Rendus* **287** (1978) 695–697
- [P<sub>1</sub>] Parimala R, Quadratic forms over polynomial rings over Dedekind domains, *Am. J. Math.* **102** (1980) 289–296
- [P<sub>2</sub>] Parimala R, Witt groups of conics, elliptic and hyperelliptic curves, *J. Number Theory* **28** (1988) 69–93
- [P<sub>3</sub>] Parimala R, Witt groups of affine three folds, *Duke Math. J.* **57** (1988) 947–954
- [P–S] Parimala R and Scharlau W, The canonical class of a curve and the extension property for quadratic forms, *Contemp. Math.* **155** (1993)
- [P–Sr] Parimala R and Sridharan R, Graded Witt ring and unramified cohomology, *K-Theory*, **6** (1992) 29–44
- [Su] Suresh V, On the canonical class of hyperelliptic curves, to appear, *Contemp. Math.* **155** (1993)

# Modular forms and differential operators

DON ZAGIER

Max-Planck-Institut für Mathematik, Gottfried-Claren Str. 26, 53225 Bonn, Germany

Dedicated to the memory of Professor K G Ramanathan

**Abstract.** In 1956, Rankin described which polynomials in the derivatives of modular forms are again modular forms, and in 1977, H Cohen defined for each  $n \geq 0$  a bilinear operation which assigns to two modular forms  $f$  and  $g$  of weight  $k$  and  $l$  a modular form  $[f, g]_n$  of weight  $k + l + 2n$ . In the present paper we study these "Rankin-Cohen brackets" from two points of view. On the one hand we give various explanations of their modularity and various algebraic relations among them by relating the modular form theory to the theories of theta series, of Jacobi forms, and of pseudodifferential operators. In a different direction, we study the abstract algebraic structure ("RC algebra") consisting of a graded vector space together with a collection of bilinear operations  $[, ]_n$  of degree  $+2n$  satisfying all of the axioms of the Rankin-Cohen brackets. Under certain hypotheses, these turn out to be equivalent to commutative graded algebras together with a derivation  $\partial$  of degree 2 and an element  $\Phi$  of degree 4, up to the equivalence relation  $(\partial, \Phi) \sim (\partial - \phi E, \Phi - \phi^2 + \partial(\phi))$  where  $\phi$  is an element of degree 2 and  $E$  is the Euler operator (= multiplication by the degree).

**Keywords.** Modular forms; Jacobi forms; pseudodifferential operators; vertex operator algebras.

The derivative of a modular form is not a modular form. Nevertheless, there are many interesting connections between differential operators and the theory of modular forms. For instance, every modular form (by which we shall always mean a holomorphic modular form in one variable of integral weight) satisfies a nonlinear third order differential equation with constant coefficients; in another direction, if such a form  $f(\tau)$  is expressed as a power series  $\varphi(t(\tau))$  in a local parameter  $t(\tau)$  which is a meromorphic modular function of  $\tau$ , then the power series  $\varphi(t)$  satisfies a linear differential equation of order  $k + 1$  with algebraic coefficients, where  $k$  is the weight of  $f$ . This latter fact, which leads to many connections between the theory of modular forms and the theory of hypergeometric and other special differential equations, played an important role in the development of both theories in the 19th century and up to the work of Fricke and Klein, but surprisingly little role in more modern investigations.

In 1956, R. A. Rankin [Ra] gave a general description of the differential operators which send modular forms to modular forms. A very interesting special case of this general setup are certain bilinear operators on the graded ring  $M_*(\Gamma)$  of modular forms on a fixed group  $\Gamma \subset PSL(2, \mathbb{R})$  which were introduced by H. Cohen [Co] and which have had many applications since then. These operators, which we call the Rankin-Cohen brackets, will be the main object of study in the present paper. On the one hand, we will be interested in understanding from various points of view

"why" these operators on modular forms have to exist. These different approaches (in particular, via Jacobi forms and via pseudodifferential operators) give different explanations and even different definitions of the operators, and although the definitions differ only by constants, the constants turn out to depend in a subtle way on the parameters involved and to lead to quite complicated combinatorial problems. On the other hand, we will try to understand what kind of an additional algebraic structure these operators give to the ring  $M_*(\Gamma)$  and what other examples of the same algebraic structure can be found in (mathematical) nature. We will give a partial structure theorem showing that the algebraic structure in question is more or less equivalent to that of a graded algebra together with a derivation of degree 2 and an element of degree 4. The results will be far from definitive, our main object being to formulate certain questions and perhaps arouse some interest in them.

## 1. The Rankin-Cohen bilinear operators

Let  $f(\tau)$  and  $g(\tau)$  denote two modular forms of weight  $k$  and  $l$  on some group  $\Gamma \subset \mathrm{PSL}(2, \mathbb{R})$ . We denote by  $D$  the differential operator  $\frac{1}{2\pi i} \frac{d}{d\tau} = q \frac{d}{dq}$  (where  $q = e^{2\pi i \tau}$  as usual) and use  $f', f'', \dots, f^{(n)}$  freely instead of  $Df, D^2f, \dots, D^n f$ . The Rankin-Cohen bracket of  $f$  and  $g$  is defined by the formula

$$[f, g]_n(\tau) = \sum_{r+s=n} (-1)^r \binom{n+k-1}{s} \binom{n+l-1}{r} f^{(r)}(\tau) g^{(s)}(\tau).$$

(The normalization here is different from that in [Co] and has been chosen so that  $[f, g]_n$  is in  $\mathbb{Z}[[q]]$  if  $f$  and  $g$  are.) The basic fact is that this is a modular form of weight  $k+l+2n$  on  $\Gamma$ , so that the graded vector space  $M_*(\Gamma)$  possesses not only the well-known structure as a commutative graded ring, corresponding to the usual bracket, but also an infinite set of further bilinear operations  $[\cdot, \cdot]_n: M_* \otimes M_* \rightarrow M_{*+2n}$ . We shall be interested in seeing what kind of an algebraic structure this is and what other examples of such a structure arise. Let us start by recalling why  $[f, g]_n$  is a modular form. There are at least two ways to see this.

The first is to associate to the modular form  $f(\tau)$  the formal power series

$$\tilde{f}(\tau, X) = \sum_{n=0}^{\infty} \frac{f^{(n)}(\tau)}{n!(n+k-1)!} (2\pi i X)^n$$

introduced by Kuznetsov [Ku] and Cohen [Co]. Then the higher brackets of  $f$  and  $g$  given by

$$\tilde{f}(\tau, -X) \tilde{g}(\tau, X) = \sum_{n=0}^{\infty} \frac{[f, g]_n(\tau)}{(n+k-1)!(n+l-1)!} (2\pi i X)^n \quad (f \in M_k, g \in M_l).$$

On the other hand,  $\tilde{f}$  satisfies the transformation law ([Ku], Theorem 1, [Co], Theorem 7.1a)

$$\tilde{f}\left(\gamma(\tau), \frac{X}{(c\tau+d)^2}\right) = (c\tau+d)^k e^{cX/(c\tau+d)} \tilde{f}(\tau, X) \quad \left(\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma, \gamma(\tau) = \frac{a\tau+b}{c\tau+d}\right)$$



Indeed, this identity is equivalent by comparison of coefficients to the sequence of identities

$$\frac{f^{(n)}(\gamma(\tau))}{n!(n+k-1)!} = \sum_{m=0}^n \frac{(2\pi ic)^{n-m} (c\tau+d)^{k+n+m}}{(n-m)!} \frac{f^{(m)}(\tau)}{m!(m+k-1)!} \quad (n \geq 0),$$

and these are easily proved by induction on  $n$ . [For a non-inductive proof of (4), observe that (2) is the unique power series solution of the differential equation  $\left(\frac{\partial}{\partial \tau} - k \frac{\partial}{\partial X} - X \frac{\partial^2}{\partial X^2}\right) \tilde{f} = 0$  with initial conditions  $\tilde{f}(\tau, 0) = (k-1)!^{-1} f(\tau)$ , and verify that  $(c\tau+d)^{-k} e^{-cX/(c\tau+d)} \tilde{f}(\gamma(\tau), X/(c\tau+d)^2)$  satisfies the same conditions.] Now identity (4) and the corresponding formula for  $\tilde{g}$  imply that the product occurring on the left-hand side of (3) is multiplied by  $(c\tau+d)^{k+l}$  under the transformation  $(\tau, X) \mapsto (\gamma(\tau), (c\tau+d)^{-2}X)$  (the exponential factors drop out because of the minus sign in (3)), and this says that the coefficient of  $X^n$  in this product transforms like a modular form of weight  $k+l+2n$  for all  $n$ . Since the holomorphy at the cusps is also easy to check, this proves the assertion.

For the second proof, which will also make it clear that the operator  $[\cdot, \cdot]_n$  is the only bilinear differential operator of degree  $2n$  sending modular forms to modular forms – a fact which can be seen in many other ways – is to look at the effect of this operator on theta series. Recall that if  $Q: \mathbb{Z}^m \rightarrow \mathbb{Z}$  is a positive definite quadratic form in  $m$  variables and  $P: \mathbb{Z}^m \rightarrow \mathbb{C}$  a spherical function of even degree  $d$  with respect to  $Q$  (i.e. a homogeneous polynomial of degree  $d$  in  $m$  variables which is annihilated by the Laplacian  $\Delta_Q$  associated to  $Q$ ), then the theta series  $f(\tau) = \Theta_{Q,P}(\tau) = \sum_{x \in \mathbb{Z}^m} P(x) q^{Q(x)}$  is a modular form of weight  $k = d + m/2$  on some subgroup  $\Gamma \subset PSL(2, \mathbb{Z})$  of finite index. If  $g = \Theta_{Q',P'}$  is a second such theta series of weight  $l = d' + m'/2$ , then the function

$$\begin{aligned} h(\tau) &= \sum_{r+s=n} c_{r,s} f^{(r)}(\tau) g^{(s)}(\tau) \\ &= \sum_{(x,x') \in \mathbb{Z}^{m+m'}} \left( P(x) P'(x') \sum_{r+s=n} c_{r,s} Q(x)^r Q'(x')^s \right) q^{Q(x)+Q'(x')} \end{aligned}$$

will be a modular form (of weight  $k+l+2n$ ) if and only if the homogeneous polynomial of weight  $d+d'+2n$  appearing in parentheses is spherical with respect to the combined Laplacian  $\Delta_Q + \Delta_{Q'}$ . But a short calculation, facilitated by choosing coordinates in which  $Q(x) = \sum_{i=1}^m x_i^2$ , shows that  $\Delta_Q(P(x)Q(x)')$  equals  $4r(r+k-1)P(x)Q(x)^{r-1}$ , so this will happen if and only if  $r(r+k-1)c_{r,s} + (s+1)(s+l)c_{r-1,s+1}$  vanishes for all  $r$  and  $s$ , i.e., if the  $c_{r,s}$  are proportional to  $(-1)^r \binom{n+k-1}{s} \binom{n+l-1}{r}$ .

## 2. Algebraic properties of the Rankin–Cohen brackets

The brackets introduced in §1 satisfy a number of algebraic identities. First, we have the obvious (anti-)commutativity property

$$[f, g]_n = (-1)^n [g, f]_n, \quad (5)$$

for all  $n$ . The 0th bracket, as already mentioned, is usual multiplication, so satisfies

the identities

$$[[f, g]_0, h]_0 = [f, [g, h]_0]_0 \quad (6)$$

making  $(M_*, [\cdot]_0)$  into a commutative and associative algebra. We also have the formulas

$$[f, 1]_0 = [1, f]_0 = f, \quad [f, 1]_n = [1, f]_n = 0 \quad (n > 0) \quad (7)$$

(because the binomial coefficient  $\binom{n-1}{n}$  in (1) is zero), which say that the unit of this algebra structure has trivial higher brackets with all of  $M_*$ . The 1st bracket, given by

$$[f, g]_1 = -[g, f]_1 = kf'g' - l'f'g \in M_{k+l+2} \quad (f \in M_k, g \in M_l),$$

satisfies the Jacobi identity

$$[[fg]_1, h]_1 + [[gh]_1, f]_1 + [[hf]_1, g]_1 = 0, \quad (8)$$

giving  $M_{*-2}$  the structure of a graded Lie algebra. (From now on, we often drop the comma in the notation for the brackets). The double brackets  $[[\cdot]_0]_1$  and  $[[\cdot]_1]_0$  satisfy the identities

$$[[fg]_0, h]_1 + [[gh]_0, f]_1 + [[hf]_0, g]_1 = 0 \quad (9)$$

and

$$m[[fg]_1, h]_0 + l[[gh]_1, f]_0 + k[[hf]_1, g]_0 = 0 \quad (f \in M_k, g \in M_l, h \in M_m) \quad (10)$$

(the first one in which the weights play a role) as well as the mixed relations

$$[[fg]_0, h]_1 = [[gh]_1, f]_0 - [[hf]_1, g]_0, \quad (11a)$$

$$(k + m + l)[[fg]_1, h]_0 = k[[hf]_0, g]_1 - l[[gh]_0, f]_1, \quad (11b)$$

the first of which says that the Lie bracket with a fixed element of  $M_*$  acts as a derivation with respect to the associative algebra structure  $[\cdot]_0$ . (A space having simultaneously the structures of an associative and a Lie algebra, with the latter acting via derivations on the former, is called a *Poisson algebra*.) The relations (6)–(11), which are not all independent, describe all identities relating the 0th and 1st brackets. At the next level, the relations involving the second bracket

$$[f, g]_2 = \binom{k+1}{2} fg'' - (k+1)(l+1)f'g' + \binom{l+1}{2} f''g \in M_{k+l+4} \\ (f \in M_k, g \in M_l)$$

are already quite complicated. Starting with  $f \otimes g \otimes h \in M_k \otimes M_l \otimes M_m$  we can already make nine trilinear expressions of weight  $k + l + m + 4$ , namely  $[[fg]_0, h]_2$ ,  $[[fg]_1, h]_1$ ,  $[[fg]_2, h]_0$  and their cyclic permutations. (The non-cyclic permutations give the same elements up to sign by (5).) The space they span has dimension 3, a basis being given by the first or the last group, which are mutually related by

$$(k+1)(l+1)[[fg]_0, h]_2 = -m(m+1)[[fg]_2, h]_0 \\ + (k+1)(k+l+1)[[gh]_2, f]_0 + (l+1)(k+l+1)[[hf]_2, g]_0, \quad (12a)$$

$$\begin{aligned} (k+l+m+1)(k+l+m+2)[[fg]_2h]_0 &= (k+1)(l+1)[[fg]_0h]_2 \\ &- (k+1)(k+l+1)[[gh]_0f]_2 - (l+1)(k+l+1)[[hf]_0g]_2, \end{aligned} \quad (12b)$$

while the second group (which is linearly dependent by virtue of the Jacobi identity (8)) is expressed in terms of these by

$$[[fg]_1h]_1 = [[gh]_0f]_2 - [[hf]_0g]_2 + [[gh]_2f]_0 - [[hf]_2g]_0. \quad (13)$$

Of course we could go on in this way, giving more and more axioms for the bracket operations of various degrees. However, it is not obvious how the whole set of relations looks, or even when we have a complete defining set for a bracket of given order. For instance, although the bracket  $[\cdot]_2$  satisfies no trilinear relations like (6) or (8), a simple dimension count shows that the permutations of the  $r$ -fold 2-brackets  $[\dots[[fg]_2h]_2\dots]_2$  are linearly dependent for all sufficiently large  $r$ , but it is not clear how far we would have to go to get the first relation or how much further to ensure that all subsequent relations obtained would be consequences of ones already found. In §3 we will give an infinite collection of trilinear relations among the Rankin-Cohen brackets which possibly may generate all relations, though we do not know this.

However, even not knowing a complete (let alone minimal) collection of universal identities satisfied by the Rankin-Cohen brackets, one can investigate the class of graded vector spaces having bracket operations which satisfy these identities and try to elucidate their structure. This will be done in §5–6. First, however, we look at two other structures on modular forms which give new explanations of the existence of the bracket operations (1) and shed further light on their algebraic nature.

### 3. Rankin-Cohen operators and Jacobi-like forms

We fix a subgroup  $\Gamma$  of  $PSL(2, \mathbb{R})$ . For each integer  $k > 0$  let  $J_k = J_k(\Gamma)$  be the set of all holomorphic functions  $\phi(\tau, X)$  on  $\mathcal{H} \times \mathbb{C}$  ( $\mathcal{H}$  = upper half-plane) satisfying

$$\phi\left(\gamma(\tau), \frac{X}{(c\tau+d)^2}\right) = (c\tau+d)^k e^{cX/(c\tau+d)} \phi(\tau, X) \quad \left(\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma\right) \quad (14)$$

(i.e., equation (4) with  $\phi$  in place of  $\tilde{f}$ ) as well as the usual holomorphy conditions at the cusps. We call the elements of  $J_k$  *Jacobi-like of weight  $k$*  because they satisfy one of the two characteristic functional equations of Jacobi forms. (The other one, which does not concern us here, involves translations of  $z$  by elements of the lattice  $\mathbb{Z}\tau + \mathbb{Z}$ , where  $X$  is proportional to  $z^2$ . See [EZ] for the theory of Jacobi forms, and in particular §3 of [EZ] for many calculations related to the ones here.)

Clearly the restriction of a Jacobi-like form to  $X=0$  is a modular form of weight  $k$  on  $\Gamma$ , and the kernel of this map  $J_k \rightarrow M_k = M_k(\Gamma)$  is just  $X$  times  $J_{k+2}(\Gamma)$ . The Kuznetsov-Cohen functional equation (4) says that we have a canonical section  $f \rightarrow (k-1)!\tilde{f}$  of  $J_k \rightarrow M_k$ , so that the sequence

$$0 \rightarrow J_{k+2}(\Gamma) \xrightarrow{\cdot X} J_k(\Gamma) \xrightarrow{X=0} M_k(\Gamma) \rightarrow 0$$

is exact and splits canonically. This implies that there is a bijection between Jacobi-like

forms of weight  $k$  and sequences of modular forms of weight  $k + 2n$  ( $n \geq 0$ ). Then the multiplication of Jacobi-like forms induces bilinear pairings  $M_* \otimes M_* \rightarrow M_{*+*+2n}$ , and these must be multiples of the Rankin-Cohen brackets. We now look at the details.

If we write  $\phi(\tau, X) \in J_k(\Gamma)$  as  $\sum_{n=0}^{\infty} \phi_n(\tau)(2\pi i X)^n$ , then comparing coefficients of  $X^n$  in the defining functional equation (14) gives the functional equations

$$(c\tau + d)^{-k-2n} \phi_n(\gamma(\tau)) = \sum_{m=0}^n \frac{1}{m!} \left( \frac{1}{2\pi i} \frac{c}{c\tau + d} \right)^m \phi_{n-m}(\tau) \quad \left( n \geq 0, \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma \right), \quad (15)$$

and conversely any sequence of holomorphic functions  $\phi_n(\tau)$  satisfying (15) and a growth condition at cusps defines an element of  $J_k(\Gamma)$ . Equations (15) are in turn equivalent to the sequence of transformation laws

$$\phi_0 \in M_k, \quad k\phi_1 - \phi'_0 \in M_{k+2}, \quad 2(k+2)(k+1)\phi_2 - 2(k+1)\phi'_1 + \phi''_0 \in M_{k+4}, \dots \quad (16)$$

and in general

$$h_n := \sum_{m=0}^n (-1)^m \frac{(2n-m+k-2)!}{m!} \phi_{n-m}^{(m)} \in M_{k+2n} \quad (n \geq 0). \quad (17)$$

This can be proved from (15) by induction on  $n$  just as (4) was proved, or alternatively deduced from (4), since a simple binomial coefficient identity lets us invert (17) to write

$$\phi_n(\tau) = \sum_{r+m=n} \frac{2m+k-1}{r!(r+2m+k-1)!} h_m^{(r)}(\tau) \quad (18)$$

or equivalently as

$$\phi(\tau, X) = \sum_{n=0}^{\infty} (2n+k-1) \tilde{h}_n(\tau, X) (2\pi i X)^n \quad (19)$$

and then the modularity of  $h_n$  follows inductively from (4) (applied to  $\tilde{h}_{n'}, n' < n$ ) and the Jacobi-like property of  $\phi$ . Equations (17) and (18) realize the afore-mentioned bijection between  $J_k(\Gamma)$  and  $\Pi_n M_{k+2n}(\Gamma)$ . Now to get the bracket operations we consider the Cohen-Kuznetsov lifts of two modular forms  $f \in M_k, g \in M_l$ . If  $\alpha$  and  $\beta$  are two complex numbers, then the product  $\phi(\tau) = \tilde{f}(\tau, \alpha X) \tilde{g}(\tau, \beta X)$  will be Jacobi-like with respect to the variable  $(\alpha + \beta)X$ , since the exponential factors in (14) multiply. The case  $\alpha + \beta = 0$  (when we can normalize to  $\beta = -\alpha = 1$ ) was the case used to obtain the Rankin-Cohen brackets in §1. If  $\alpha + \beta$  is different from 0, we can normalize it to be equal to 1 by rescaling  $X$ . Then  $\phi$  belongs to  $J_{k+l}$  and has an expansion of the form (19). The coefficient of  $(2\pi i X)^n$  in  $\phi$  is given by

$$\phi_n(\tau) = \sum_{r+s=n} \frac{\alpha^r \beta^s}{r!s!(r+k-1)!(s+l-1)!} f^{(r)}(\tau) g^{(s)}(\tau),$$

so by Leibniz's rule the modular forms  $h_n$  defined by (17) (with  $k+l$  in place of  $k$ )

are given by

$$h_n(\tau) = \sum_{p+q+r+s=n} \frac{(2n-p-q+k+l-2)!}{p!q!r!s!(r+k-1)!(s+l-1)!} \alpha^r \beta^s f^{(p+r)}(\tau) g^{(q+s)}(\tau) \quad (n \geq 0), \quad (20)$$

This is a combination of products of derivatives of  $f$  and  $g$  which is modular of weight  $k+l+2n$  and hence must be a multiple  $\kappa_n = \kappa_n(k, l; \alpha, \beta)$  of the Rankin-Cohen bracket  $[f, g]_n$ , so as  $\alpha$  and  $\beta = 1 - \alpha$  vary we get infinitely many explanations of the existence of these brackets. Our next job is to compute the scalar  $\kappa_n$ .

We define for each  $n$  a polynomial  $H_n$  of four variables, of degree  $n$  in the first two and homogeneous of degree  $n$  in the last two, by

$$H_n(k, l; X, Y) = \sum_{r+s=n} (-1)^r \binom{n+k-1}{s} \binom{n+l-1}{r} X^r Y^s, \quad (21)$$

so that equation (1) can be rewritten as

$$[f, g]_n = H_n(k, l; D_{\tau_1}, D_{\tau_2})(f(\tau_1)g(\tau_2))|_{\tau_1=\tau_2=\tau} \quad (f \in M_k, g \in M_l). \quad (22)$$

The polynomials  $H_n$ , whose definition can also be written

$$H_n(k, l; X, Y) = \frac{1}{n!} \left( -X \frac{\partial}{\partial \eta} + Y \frac{\partial}{\partial \xi} \right)^n (\xi^{k+n-1} \eta^{l+n-1}) \Big|_{\xi=\eta=1},$$

satisfy many algebraic identities. We mention in particular

$$H_n(k, l; X, Y) = (-1)^n H_n(l, k; Y, X), \quad (23)$$

$$H_n(k, l; X, Y) = H_n(l, -k-l-2n+2; Y, -X-Y), \quad (24)$$

$$\begin{aligned} & \frac{n!(n+k+l-2)!}{(n+k-1)!(n+l-1)!} H_n(k, l; X, Y) H_n(k, l; \alpha, \beta) \\ &= \sum_{r+s+t=n} \frac{(2r+2s+t+k+l-2)!}{r!s!t!(r+k-1)!(s+l-1)!} (\alpha X)^r (\beta Y)^s (-\gamma Z)^t \\ & \quad (\alpha + \beta + \gamma = X + Y + Z = 0). \end{aligned} \quad (25)$$

The first two of these say that the 6-argument function

$$\begin{bmatrix} X & Y & Z \\ k & l & m \end{bmatrix} = H_n(k, l; X, Y) \quad (k+l+m=2-2n, \quad X+Y+Z=0)$$

is symmetric under even and  $(-1)^n$ -symmetric under odd permutations of its three columns, and the third (for  $k+l \notin \mathbb{Z}$ ) can then be rewritten more symmetrically as

$$\begin{aligned} & \sum_{r+s+t=n} \frac{(\alpha X)^r (\beta Y)^s (\gamma Z)^t}{r!s!t!(r+k-1)!(s+l-1)!(t+m-1)!} \\ &= \frac{1}{(n+k-1)!(n+l-1)!(n+m-1)!} \begin{bmatrix} X & Y & Z \\ k & l & m \end{bmatrix} \begin{bmatrix} \alpha & \beta & \gamma \\ k & l & m \end{bmatrix} \end{aligned}$$

where  $x!$  denotes  $\Gamma(x+1)$  for  $x \notin \mathbb{Z}$ .

identity (25) is trivial. To prove (24) and (26) we observe that  $H_n(k, l; \alpha, \beta, \gamma)$  satisfies the differential equations

$$\left(X \frac{\partial}{\partial X} + Y \frac{\partial}{\partial Y}\right) H_n = 0, \quad \left(k \frac{\partial}{\partial X} + \frac{\partial^2}{\partial X^2} + l \frac{\partial}{\partial Y} + \frac{\partial^2}{\partial Y^2}\right) H_n = 0 \quad (26)$$

(the first is Euler's equation saying that  $H_n$  is homogeneous of degree  $n$  in  $X$  and  $Y$ , and the second was already used implicitly in the proof of modularity of  $[\Theta_{p,q}, \Theta_{p',q'}]_n$  in § 1), and these characterize  $H_n$  uniquely up to a scalar factor as a function of  $X$  and  $Y$ . Thus to prove (24) we verify, using (26), that the right-hand side also satisfies (26), and then fix the normalizations by taking  $Y = 0$  and using (23). Similarly, to prove (25) we verify that the expression on the right satisfies (26) and hence is a multiple (depending on  $\alpha$  and  $\beta$ ) of  $H_n(k, l; X, Y)$ ; by the symmetry in  $(X, Y, Z)$  and  $(\alpha, \beta, \gamma)$ , this multiple must be a scalar multiple  $\lambda_n(k, l)$  of  $H_n(k, l; \alpha, \beta)$ , and the value of  $\lambda_n(k, l)$  is fixed by specializing to  $\alpha = Y = 0$ . One can also prove both identities using generating functions; for instance, we have

$$\begin{aligned} \sum_{n=0}^{\infty} H_n(k-n+1, l; X, Y) T^n &= \sum_{r,s \geq 0} (-1)^r \binom{k}{s} \binom{r+s+l-1}{r} X^r Y^s \\ &= \sum_{s \geq 0} \binom{k}{s} (TY)^s (1+TX)^{-l-s} = \frac{(1-TZ)^k}{(1+TX)^{k+l}} \quad (Z = -X - Y) \end{aligned}$$

and hence  $H_n(k-n+1, l; X, Y) = (-1)^n H_n(-k-l-n+1, l; Z, Y)$ , which is equivalent to (24).

Now returning to (20), where  $\alpha + \beta = 1$ , we see from (25) and (21) that

$$h_n(\tau) = \frac{n!(n+k+l-2)!}{(n+k-1)!(n+l-1)!} H_n(k, l; \alpha, \beta) [f, g]_n(\tau).$$

(This actually gives another proof of (25), since we already knew that  $h_n$  had to be a multiple of  $[f, g]_n$ , so the right-hand side of (25) must be a multiple of  $H_n(k, l; X, Y)$ .) Changing to the inhomogeneous notation, we can summarize what we have proved as:

### PROPOSITION

For  $f \in M_k(\Gamma)$ ,  $g \in M_l(\Gamma)$  ( $k, l > 0$ ) and  $\alpha, \beta \in \mathbb{C}$  we have the identity

$$\tilde{f}(\tau, \alpha X) \tilde{g}(\tau, \beta X) = \sum_{n=0}^{\infty} c_n(k, l; \alpha, \beta) [\tilde{f}, \tilde{g}]_n(\tau, (\alpha + \beta)X) (2\pi i X)^n \quad (27)$$

with

$$c_n = (2n+k-1) \frac{n!(n+k+l-2)!}{(n+k-1)!(n+l-1)!} H_n(k, l; \alpha, \beta). \quad (28)$$

Applying this proposition twice, we find that, if  $h \in M_m(\Gamma)$  is a third modular form on  $\Gamma$ , then

$$\begin{aligned} \tilde{f}(\tau, \alpha X) \tilde{g}(\tau, \beta X) \tilde{h}(\tau, \gamma X) &= \\ \sum_{n,p \geq 0} c_n(k, l; \alpha, \beta) c_p(k+l+2n, m; \alpha + \beta, \gamma) \tilde{F}_{n,p}(\tau, (\alpha + \beta + \gamma)X) (2\pi i X)^{n+p} \end{aligned}$$

with  $F_{n,p}(\tau) = [[f, g]_n h]_p$ . Since the expression on the left is symmetric in its arguments, we get:

## COROLLARY

For  $f \in M_k(\Gamma)$ ,  $g \in M_l(\Gamma)$  and  $h \in M_m(\Gamma)$  and  $\alpha, \beta, \gamma \in \mathbb{C}$ , the expression

$$\sum_{n=0}^r c_n(k, l; \alpha, \beta) c_{r-n}(k+l+2n, m; \alpha+\beta, \gamma) [[f, g]_n, h]_{r-n} \in M_{k+l+m+2r}(\Gamma) \\ (r \in \mathbb{Z}_{\geq 0}),$$

with  $c_n$  given by (28), is symmetric under all permutations of  $(f, k, \alpha)$ ,  $(g, l, \beta)$ ,  $(h, m, \gamma)$ .

Varying  $r$  and comparing coefficients of the various monomials in  $\alpha$ ,  $\beta$  and  $\gamma$ , we systematically obtain in this way universal identities satisfied by the Rankin-Cohen brackets of the sort studied in §2. For instance, the triple brackets  $[[fg]_* h]_*$  can always be expressed (in general, in many ways) as linear combinations of the triple brackets  $[[fh]_* g]_*$ .

Finally, we mention that combinational identities similar to (24) and (25) occur, in a somewhat related context, in the paper [IZ].

## 4. The Rankin-Cohen operators and pseudodifferential operators

This connection was suggested to the author by Yu. Manin and will be treated in detail in the joint paper [MZ], so we give only a few indications.

Let  $D$  as before be the differential operator  $(2\pi i)^{-1} d/d\tau$ . (The factor  $2\pi i$ , introduced earlier for convenience, is more of a nuisance now, but we will let it be.) Then by a formal pseudodifferential operator we mean a formal power series  $\sum_{n=0}^{\infty} g_n(\tau) D^{-n}$  where the  $g_n$  are holomorphic functions in the upper half-plane. We can multiply two such series by Leibniz's rule

$$\left( \sum_{m=0}^{\infty} f_m(\tau) D^{-m} \right) \left( \sum_{n=0}^{\infty} g_n(\tau) D^{-n} \right) = \sum_{m,r,n \geq 0} \binom{-m}{r} f_m(\tau) g_n^{(r)}(\tau) D^{-m-r-n},$$

and the pseudodifferential operators in this way form an associative, but of course not commutative, ring.

Now if we consider some modular group  $\Gamma$  acting on the upper half-plane, then  $\Gamma$  also acts on  $D$  via  $D \mapsto (\tau + d)^2 D$ , so it makes sense to speak of a pseudodifferential operator  $\sum_{n=0}^{\infty} g_n(\tau) D^{-n}$  being  $\Gamma$ -invariant. If  $n_0 > 0$  is the smallest index with  $g_{n_0} \neq 0$  for such an operator, then it is easily seen that  $g_{n_0}$  is a modular form of weight  $k = 2n_0$ ,  $g_{n_0+1} + \frac{1}{2}(n_0+1)g'_{n_0}$  is a modular form of weight  $k+2$ , etc. This is reminiscent of the equations (15), and indeed, a calculation shows that the power series

$$\sum_{n=n_0}^{\infty} \frac{g_n(\tau)}{n!(n-1)!} (-2\pi i X)^{n-n_0}$$

belongs to  $J_k(\Gamma)$ , setting up a 1:1 correspondence between invariant pseudodifferential operators of the form  $\sum_{n \geq n_0} g_n D^{-n}$  and Jacobi-like forms of weight  $k$ . Combining this

with the Kuznetsov-Cohen lifting (4), we find that there is a canonical lifting

$$f(\tau) \mapsto \mathcal{D}[f] = \sum_{r=0}^{\infty} (-1)^r \frac{(r+k/2)!(r+k/2-1)!}{r!(r+k-1)!} f^{(r)}(\tau) D^{-r-k/2} \\ (f \in M_k, k > 0 \text{ even})$$

from modular forms to pseudodifferential operators, and that conversely any  $\Gamma$ -invariant pseudodifferential operator can be expanded as a sum of such lifts. In particular, since the product of two  $\Gamma$ -invariant pseudodifferential operators is another one, we can associate to two modular forms  $f \in M_k, g \in M_l$  a sequence of modular forms  $\{h_n\}_{n \geq 0}$  via

$$\mathcal{D}[f] \cdot \mathcal{D}[g] = \sum_{n=0}^{\infty} \mathcal{D}[h_n] \quad (h_n \in M_{k+l+2n}).$$

Then, just as in §3, the uniqueness of the Rankin-Cohen brackets implies that  $h_n$  must be some universal factor  $t_n = t_n(k, l)$  of  $[f, g]_n$ . Since, unlike the situation in §3 where the definition of the modular forms  $h_n$  depended on an arbitrary parameter  $\alpha$ , the present operation is completely canonical, one would expect the scalar factor occurring to be very simple. Surprisingly, it is not: the combinatorial calculations needed here are far worse than the already complicated ones in §3. The formula for  $t_n(k, l)$ , as well as other aspects of the connection between pseudodifferential operators and modular forms (including a connection with super-pseudodifferential operators in the case of modular forms of odd weight), will be discussed in [MZ].

## 5. Definition and examples of Rankin-Cohen algebras

We define a *Rankin-Cohen algebra* (or RC algebra for short) over a field  $K$  as a graded  $K$ -vector space  $M_* = \bigoplus_{k \geq 0} M_k$  (with  $M_0 = K \cdot 1$  and  $\dim_K M_k$  finite for all  $k$ )

together with bilinear operations  $[\cdot, \cdot]_n: M_k \otimes M_l \rightarrow M_{k+l+2n}$  ( $k, l, n \geq 0$ ) which satisfy (5)–(13) and all the other algebraic identities satisfied by the Rankin-Cohen brackets. In view of the remarks at the end of §2, this may seem like a strange definition, since we do not know how to give a complete set of axioms. Nevertheless, we will be able to construct examples and, to a large extent, to clarify the structure of these objects. The situation should be thought of as analogous to building up the theory of Lie algebras starting with the observation that the operation  $[X, Y] = XY - YX$  in an associative algebra seems to have interesting properties. One could then define Lie algebras as algebras with a bracket satisfying all algebraic identities universally satisfied by this standard bracket in any associative algebra, and a good many results could be proved without knowing a complete generating set for these identities. One would initially be forced to look at subspaces of associative algebras closed under the standard bracket, but would eventually prove that all Lie algebras arise this way (existence of the universal enveloping algebra) and also that all universal identities satisfied by the bracket are in fact consequences of anticommutativity and the Jacobi identity. In the same way, we will start by considering RC algebras which are subspaces closed under all bracket operations of some standard examples, and then show that, under some general hypotheses, all RC algebras in fact arise in this way.



We will suppose the ground field  $K$  to be of characteristic 0 (in our examples it is usually  $\mathbb{Q}$  or  $\mathbb{C}$ ) although it is clear that the theory makes sense in any characteristic or, for that matter, even if we work over  $\mathbb{Z}$  rather than a field.

*Example 1.* Since modular forms and their derivatives do not satisfy any universal relation, the only identities satisfied by the Rankin-Cohen brackets on  $M_*(\Gamma)$  are those following from the formula (1) and Leibniz's rule. The basic example of an RC algebra is therefore given by

*Definition.* Let  $R_*$  be a commutative graded algebra with unit over  $K$  together with a derivation  $D: R_* \rightarrow R_*$  of degree 2 (i.e.  $D(R_k) \subseteq R_{k+2}$  for all  $k$  and  $(fg)' = f'g + fg'$ , where as before  $f', f'', \dots, f^{(r)}$  denote  $Df, D^2f, \dots, D^r f$ ), and define  $[\cdot, \cdot]_{D,n}$  by

$$[f, g]_{D,n} = \sum_{r+s=n} (-1)^r \binom{n+k-1}{s} \binom{n+l-1}{r} f^{(r)} g^{(s)} \in R_{k+l+2n} \quad (f \in R_k, g \in R_l). \quad (29)$$

Then  $(R_*, [\cdot, \cdot]_{D,n})$  is an RC algebra which we will call the *standard RC algebra on  $(R_*, D)$* .

Since a subspace of an RC algebra which contains 1 and is closed under all the bracket operations is obviously again an RC algebra, this gives us a large further class of examples, the sub-RC algebras of the standard ones. A basic question (the analogue of the question of the existence of universal enveloping algebras in the Lie algebra case) is whether every RC algebra can in fact be realized in this way. We will give an affirmative answer under a weak additional hypothesis below.

*Example 2.* Our original example of an RC algebra,  $M_*(\Gamma)$  with the brackets defined by (1), is not a standard algebra, since  $M_*(\Gamma)$  is not closed under  $D = (2\pi i)^{-1} d/d\tau$ . Of course it is a subalgebra of a standard RC-algebra in a variety of ways, since we can take  $R_*$  to be any algebra of functions on the upper half-plane which contains  $M_*(\Gamma)$  and is closed under differentiation (e.g. the space of all  $C^\infty$  or of all holomorphic functions). However, we would like an  $R_*$  which is not too big. Let us look in more detail at the case  $\Gamma = PSL(2, \mathbb{Z})$ . Here  $M_*(\Gamma) = \mathbb{C}[Q, R]$ , where  $Q = 1 + 240q + \dots$  and  $R = 1 - 504q - \dots$  are the normalized Eisenstein series of weights 4 and 6 (in Ramanujan's notation). As is well-known, their derivatives are given by  $Q' = \frac{1}{3}(PQ - R)$  and  $R' = \frac{1}{2}(PR - Q^2)$ , where  $P = 1 - 24q - \dots$  is the normalized Eisenstein series of weight 2, and since we also have  $P' = \frac{1}{12}(P^2 - Q)$  this says that  $M_*(\Gamma)$  is contained in the standard RC algebra on  $(\mathbb{C}[P, Q, R], D)$ . Now, forgetting modular forms and the interpretation of  $P, Q$  and  $R$  as functions, we can express this example purely algebraically: let  $K$  be a field of characteristic 0 and define a derivation on the polynomial algebra over  $K$  on three graded generators  $P, Q, R$  of degrees 2, 4 and 6 by

$$D = \frac{P^2 - Q}{12} \frac{\partial}{\partial P} + \frac{PQ - R}{3} \frac{\partial}{\partial Q} + \frac{PR - Q^2}{2} \frac{\partial}{\partial R}: K[P, Q, R]_* \rightarrow K[P, Q, R]_{*+2}; \quad (30)$$

then the subalgebra generated by  $Q$  and  $R$  is closed under the bracket operators  $[\cdot, \cdot]_n = [\cdot, \cdot]_{D,n}$  defined by (29) for all  $n \geq 0$ . From an algebraic point of view this is not

at all obvious (except for  $n=0$ ), although one can easily check a few examples, e.g. (with  $1728\Delta = Q^3 - R^2$ )

$$\begin{aligned} [Q, R]_1 &= -3456\Delta, & [Q, \Delta]_1 &= 4R\Delta, & [R, \Delta]_1 &= 6Q^2\Delta, \\ [Q, Q]_2 &= 4800\Delta, & [Q, R]_2 &= 0, & [R, R]_2 &= -21168Q\Delta, \\ [\Delta, \Delta]_2 &= -13Q\Delta^2. \end{aligned} \quad (31)$$

*Example 3.* We try to understand the last example by observing that we also have a derivation  $\partial$  of degree 2 on the subalgebra  $M_* = K[Q, R]$  of  $R_* = K[P, Q, R]$ , defined in terms of  $D$  by

$$\partial f = Df - \frac{k}{12} Pf \in M_{k+2} \quad (f \in M_k) \quad (32)$$

or directly by

$$\partial = -\frac{R}{3} \frac{\partial}{\partial Q} - \frac{Q^2}{2} \frac{\partial}{\partial R} : M_* \rightarrow M_{*+2} \quad (33)$$

(this is a well-known fact about derivatives of modular forms, but is also clear algebraically from (30)). Of course the standard RC algebra structure on  $M_*$  associated to  $\partial$  is completely different from the one inherited from  $(R_*, D)$ . But we now see that we can reconstruct  $(R_*, \partial)$  from  $(R_*, D)$  by using (32) to define  $Df$  for  $f \in M_k$  and defining  $D(P)$  as  $\frac{1}{12}(P^2 - Q)$ . We generalize this example in the following result.

#### PROPOSITION 1

Let  $M_*$  be a commutative and associative graded  $K$ -algebra with  $M_0 = K.1$  together with a derivation  $\partial : M_* \rightarrow M_{*+2}$  of degree 2, and let  $\Phi \in M_4$ . Define brackets  $[ \ ]_{\partial, \Phi, n}$  ( $n \geq 0$ ) on  $M_*$  by

$$\begin{aligned} [f, g]_{\partial, \Phi, n} &= \sum_{r+s=n} (-1)^r \binom{n+k-1}{s} \binom{n+l-1}{r} f_r g_s \in M_{k+l+2n} \\ &\quad (f \in M_k, g \in M_l) \end{aligned} \quad (34)$$

where  $f_r \in M_{k+2r}, g_s \in M_{l+2s}$  ( $r, s \geq 0$ ) are defined recursively by

$$f_{r+1} = \partial f_r + r(r+k-1)\Phi f_{r-1}, \quad g_{s+1} = \partial g_s + s(s+l-1)\Phi g_{s-1} \quad (r, s \geq 0) \quad (35)$$

with initial conditions  $f_0 = f, g_0 = g$  (so  $f_1 = \partial f, f_2 = \partial^2 f + k\Phi f$  and similarly for  $g_s$ ). Then  $(M_*, [ \ ]_{\partial, \Phi, *})$  is an RC algebra.

*Definition.* An RC algebra will be called *canonical* if its brackets are given as in Proposition 1 for some derivation  $\partial : M_* \rightarrow M_{*+2}$  of degree +2 and some element  $\Phi \in M_4$ .

*Proof.* As already observed, our only way to verify that something is an RC algebra is to embed it into a standard RC algebra  $(R_*, [ \ ]_{D, *})$  for some larger graded ring  $R_*$  with derivation  $D$ . We take  $R_* = M[\phi]_* := M_* \otimes_K K[\phi]$ , where  $\phi$  has degree 2, and define  $D$  by

$$D(f) = \partial(f) + k\phi f \in R_{k+2} \quad (f \in M_k), \quad D(\phi) = \Phi + \phi^2 \in R_4. \quad (36)$$

(This defines  $D$  on generators of  $R_*$ , and we extend  $D$  uniquely as a derivation.) If we show that  $[f, g]_{D,n} = [f, g]_{\partial, \Phi, n}$  for  $f$  and  $g$  in  $M_*$  then we are done, since  $M_*$  is obviously closed under the brackets  $[\cdot]_{\partial, \Phi, n}$ . To this end, we observe that the brackets  $[\cdot]_{D,n}$ , just as in § 1, can be described by the generating function

$$\sum_{n=0}^{\infty} \frac{[f, g]_{D,n}}{(n+k-1)!(n+l-1)!} X^n = \tilde{f}(-X) \tilde{g}(X) \in R_*[[X]] \quad (f \in R_k, g \in R_l)$$

where

$$\tilde{f}(X) = \sum_{n=0}^{\infty} \frac{f^{(n)}}{n!(n+k-1)!} X^n, \quad \tilde{g}(X) = \sum_{n=0}^{\infty} \frac{g^{(n)}}{n!(n+l-1)!} X^n.$$

(These make sense only for  $k$  and  $l$  strictly positive, but since  $M_0 = K.1$  and the brackets (34) clearly satisfy (7), there is no harm in assuming this.) We claim that

$$e^{-\phi X} \tilde{f}(X) = \sum_{r=0}^{\infty} \frac{f_r}{r!(r+k-1)!} X^r \quad (37)$$

with  $f_r$  defined by (35), and similarly of course for  $g$ ; the assertion follows immediately since the exponential terms  $e^{\pm \phi X}$  drop out in the product  $\tilde{f}(-X) \tilde{g}(X)$ .

To prove (37), we define  $f_r$  by the generating function (37) and prove the recursion (35) by induction (the initial condition  $f_0 = f$  is obvious). Clearly (37) is equivalent to the closed formula

$$f_r = \sum_{n=0}^r \frac{(-1)^n r!(r+k-1)!}{n!(n+k-1)!(r-n)!} \phi^{r-n} f^{(n)} \in R_{k+2r}.$$

Assume inductively that we have proved that  $f_r \in M_{k+2r}$  for some  $r$ . Then

$$\begin{aligned} \partial f_r &= f'_r - (k+2r) \phi f_r \\ &= \sum_{n=0}^r \frac{(-1)^n r!(r+k-1)!}{n!(n+k-1)!(r-n)!} [\phi^{r-n} f^{(n+1)} + (r-n) \phi^{r-n-1} (\phi^2 + \Phi) f^{(n)} \\ &\quad - (k+2r) \phi^{r-n+1} f^{(n)}] \\ &= \sum_{n=0}^{r+1} \frac{(-1)^{r+1-n} r!(r+k-1)!}{n!(n+k-1)!(r+1-n)!} [n(n+k-1) - (r-n)(r+1-n) \\ &\quad + (k+2r)(r+1-n)] \phi^{r-n+1} f^{(n)} \\ &\quad + \Phi \sum_{n=0}^{r-1} \frac{(-1)^{r-n} r!(r+k-1)!}{n!(n+k-1)!(r-n-1)!} \phi^{r-n-1} f^{(n)} \\ &= f_{r+1} - r(r+k-1) \Phi f_{r-1} \in M_{k+2r+2}, \end{aligned}$$

and this simultaneously proves the recursion (35) and the inductively used assumption  $f_r \in M_{k+2r}$ .  $\square$

We observe that the definition (36) is motivated by (32) in the special case  $M_* = K[Q, R]$ ,  $\phi = \frac{1}{12} P$ ,  $\Phi = \frac{-1}{144} Q$ , and that the proof just given is merely the algebraic abstraction of the proposition on page 94 of [VZ] in that case (compare also iv) and v) on the following page for the case when  $M_*$  is the ring of modular forms on some group  $\Gamma$  other than  $PSL(2, \mathbb{Z})$ ).

## 6. A structure theorem for Rankin-Cohen algebras

A priori one would not expect that a subring of a ring  $R_*$  with derivative would ever be closed under all the infinitely many bracket operations  $[\cdot]_{D,n}$  unless it were already closed under  $D$ . The only non-trivial example which we had where this happened, the rings of modular forms on subgroups of  $PSL(2, \mathbb{R})$ , has just been explained by the construction given in the Proposition above. It is then natural to expect that this construction may suffice to yield all examples of RC algebras. In this section we will show that this is "almost" true, and write down conditions under which it is exactly true.

We therefore assume given an RC algebra  $M_*$  over a field  $K$ , and want to realize its brackets as the brackets  $[\cdot]_{\partial, \Phi, n}$  for some derivation  $\partial$  of degree 2 and some element  $\Phi$  of degree 4. Since the 0th bracket makes  $M_*$  into an ordinary commutative algebra (by virtue of equations (5)–(7)), we already have a ring structure, which we will denote from now on in the usual way by juxtaposition (i.e.  $fg$  instead of  $[f, g]_0$ ). Let us assume that this ring is an integral domain, or at least that there is one homogeneous element  $F$  of some positive degree  $N$  which is not a zero-divisor, and let  $\hat{M}_*$  be the quotient field of  $M_*$  or the ring  $M[1/F]_*$ , respectively. (It has elements of positive and negative grading and hence is not quite the kind of object considered up to now. The compatibility of all the brackets in the case of a standard RC algebra now implies that we can canonically extend the bracket operations to  $\hat{M}_*$ . For instance, the first equation in (11), which says that the Lie bracket  $[\cdot, h]_1$  with a fixed element  $h$  acts as a derivation with respect to the ring structure, forces us to define  $[f/F, h]_1$  as  $[f, h]_1/F - f[F, h]_1/F^2$ . We now define a derivation  $\partial: \hat{M}_* \rightarrow \hat{M}_{*+2}$  and an element  $\Phi \in \hat{M}_4$  by

$$\partial(f) = \frac{[F, f]_1}{NF} \quad (f \in \hat{M}_*), \quad \Phi = \frac{[F, F]_2}{N^2(N+1)F^2} \quad (38)$$

We claim that the brackets  $[\cdot]_{\partial, \Phi, n}$  associated to  $\partial$  and  $\Phi$  agree with the given brackets on  $\hat{M}_*$ . Indeed, since all formal identities among brackets which are satisfied by standard RC algebras are by definition satisfied in all RC algebras, it is enough to check this for  $(M_*, [\cdot]_*)$  a subalgebra of a standard RC algebra  $(R_*, [\cdot]_{D,*})$ . The bracket  $[\cdot]_{D,*}$  extends to  $\hat{R}_* = R_* \otimes_{M_*} \hat{M}_*$  for the same reason as before. Define  $\phi \in \hat{R}_2$  by

$$\phi = \frac{F'}{NF} \quad (F' = D(F) \in R_{N+2}).$$

Then for  $f \in R_k$  we have

$$D(f) - k\phi f - \partial(f) = \frac{1}{NF}(NFf' - k\phi f' - [F, f]_1) = 0$$

by (29) with  $n = 1$  and

$$D(\phi) - \phi^2 - \Phi = \left( \frac{F'}{NF} \right)' - \left( \frac{F'}{NF} \right)^2 - \frac{N(N+1)FF'' - (N+1)^2 F'^2}{N^2(N+1)F^2} = 0$$

by (29) with  $n=2$ , so  $D$  and  $\partial$  are indeed related by (36) and consequently  $[\cdot]_* = [\cdot]_{D,*}|_{M_*} = [\cdot]_{\partial,\Phi,*}$  by the calculation already given in the proof of Proposition 1. This shows that any RC algebra  $M_*$  which contains at least one homogeneous element  $F$  of positive degree which is not a zero-divisor is a subalgebra of a canonical RC algebra (namely,  $(\hat{M}_*, [\cdot]_{\partial,\Phi,*})$  with  $\hat{M}_* = M_*[1/F]$  and  $\partial, \Phi$  given by (38)) and hence also a sub RC algebra of a standard algebra (namely  $(\hat{M}_*[\phi], [\cdot]_{D,*})$  with  $\phi$  of degree 2 and  $D: \hat{M}[\phi]_* \rightarrow \hat{M}[\phi]_{*+2}$  defined by (36)). Note that if  $M_*$  is already embedded as a sub RC algebra of a standard RC algebra  $(R_*, [\cdot]_{D,*})$ , then this embedding extends to an embedding of  $\hat{M}_*[\phi]$  into  $\hat{R}_* = R[1/F]_*$  by  $\phi \mapsto D(F)/NF$  and this extension is compatible with the differentials by the calculations just done. We state the special case when  $M_*$  is closed under  $\partial$  and contains  $\Phi$  as:

## PROPOSITION 2

Let  $M_*$  be an RC algebra and suppose that  $M_*$  contains a homogeneous element  $F$  of some degree  $N > 0$  such that

- (i)  $F$  is not a zero-divisor;
- (ii)  $[F, M_*]_1 \subseteq (F) = M_* \cdot F$ ;
- (iii)  $[F, F]_2 \in (F^2)$ .

Then  $[\cdot]_* = [\cdot]_{\partial,\Phi,n}$  for  $\partial: M_* \rightarrow M_{*+2}$  and  $\Phi \in M_4$  as in (38), so  $M_*$  is a canonical RC algebra.

Examples of RC algebras which satisfy the conditions of Proposition 2 are the rings of modular forms  $M_*(\Gamma)$  with  $\Gamma \subset PSL(2, \mathbb{R})$  commensurable with  $\Gamma_1 = PSL(2, \mathbb{Z})$  and the RC bracket defined by (1). Indeed, on such a group we can define a modular form  $F(\tau) = \prod |\Delta|_{12}\gamma(\tau)$ , where  $\gamma$  runs over the set of left cosets  $(\Gamma \cap \Gamma_1) \backslash \Gamma_1$  and  $f|_k \gamma(\tau) = (c\tau + d)^k f(\gamma\tau)$  for  $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$  as usual. This is a modular form of weight  $N = 12[\Gamma_1 : \Gamma_1 \cap \Gamma]$  which has no zeros in the upper half-plane. Thus  $[F, f]_1/F$  ( $f \in M_k(\Gamma)$ ) and  $[F, F]_2/F^2$  are certainly holomorphic in the upper half-plane, and of course they transform with respect to  $\Gamma_1$  like modular forms (of weights  $k+2$  and 4, respectively). To see that they actually belong to  $M_*(\Gamma)$ , we must check that they are holomorphic at the cusps, but this is clear because it is obvious from (1) that  $\text{ord}_\infty([f, g]_n) \geq \text{ord}_\infty(f) + \text{ord}_\infty(g)$  for any  $f, g \in M_*(\Gamma_1)$  and the identity  $[f|_k \gamma, g|_l \gamma]_n = [f, g]_n|_{k+l+2n}\gamma$  shows that the same inequality is true at any cusp.

I do not know whether it is true that any RC algebra which is finitely generated (over the ground field  $K$ ) and an integral domain satisfies the conditions of Proposition 2 for some  $F$ . (The stated hypotheses are definitely necessary). But even apart from this, Proposition 2 is not really a satisfactory characterization, since there is no obvious way to pick  $F$ , which a priori could have arbitrarily high degree. The following sharpening of Proposition 2 gives a criterion for an RC algebra to be canonical which can be checked in a finite amount of time.

**Theorem.** Let  $(M_*, [\cdot]_*)$  be an RC algebra which is finitely generated over a field of characteristic 0. Then the following are equivalent:

- (a)  $(M_*, [\cdot]_*)$  is canonical.
- (b) for every homogeneous element  $F \in M_*$  there is an element  $G \in M_{*+2}$  such that

- (i)  $[F, f]_1 \equiv kfG \pmod{F}$  for all  $k \geq 0$  and all  $f \in M_k$ .  
(ii)  $[F, F]_2 \equiv (N+1)G^2 - (N+1)[F, G]_1 \pmod{F^2}$ .  
(c) Property (b) holds for some homogeneous  $F \in M_*$  which is not a divisor of zero.

Specifically, if  $(F, G)$  are a pair of elements satisfying (i) and (ii), and with  $F \in M_N$  not a divisor of zero, then the bracket on  $M_*$  agrees with the canonical bracket associated to

$$\partial_{F,G}(f) := \frac{[F, f]_1 - kfG}{NF} \quad (f \in M_k), \quad \Phi_{F,G} := \frac{[F, F]_2 + (N+1)([F, G]_1 - G^2)}{N^2(N+1)F^2}. \quad (39)$$

*Remarks.* The special case when  $F$  can be chosen in (c) with  $G = 0$  is Proposition 2, but because of our freedom to pick any element (homogeneous and not a divisor of 0) to verify (c), we now get an effective criterion to check whether a given RC algebra is canonical. Indeed, pick any  $F$ , say of weight  $N$ , and check whether the elements  $[F, f_i]$  are proportional to  $k_i f_i$  modulo the ideal  $(F)$ , where  $f_i$  ( $i = 1, \dots, I$ ) are homogeneous generators of  $M_*$  of weight  $k_i$ . If this is not the case, then  $M_*$  is not canonical by the implication (a)  $\Rightarrow$  (b). If it is, then pick an element  $G$  to  $M_{n+2}$  satisfying (i) and verify whether (ii) is true. If it is, then  $M_*$  is canonical by the implication (c)  $\Rightarrow$  (a) of the theorem. If it is not, then  $M_*$  is not canonical, because of the implication (a)  $\Rightarrow$  (b) and the fact that the truth of (ii) is independent of the choice of  $G$ . (Any two choices differ by a multiple of  $F$ , and if  $G_1 = G + F\phi$  with  $\phi \in M_2$  then  $[F, G_1]_1 - G_1^2 - [F, G]_1 + G^2 = F([F, \phi]_1 - 2G\phi) - \phi^2 F^2$  belongs to  $(F^2)$  by the defining property of  $G$ .)

*Example.* Let  $M_* = M_*(PSL(2, \mathbb{Z})) = \mathbb{C}[Q, R]$  with the original Rankin-Cohen bracket (1). Of course we already know that this satisfies the conditions of Proposition 2 with  $F = \Delta = \frac{Q^3 - R^2}{1728}$ , giving  $M_*$  the canonical structure associated to the derivation (33)

and the element  $\Phi = -Q/144$ . But suppose that we had not noticed this nice element  $\Delta$  and instead wanted to check the canonicalness of  $M_*$  starting with  $F = Q$ , the homogeneous element of lowest positive weight in  $M_*$ . According to the theorem, we must find an element  $G \in M_6$  satisfying (i) and (ii). Since  $M_*$  has only two generators  $Q$  and  $R$ , and by the derivation property of  $[\cdot]_1$ , it is enough to check (i) for  $f = Q$  and  $f = R$ . Using (31) we find

$$[F, Q]_1 = 0 \equiv 4Q \cdot \frac{R}{3} \pmod{Q}, \quad [F, R]_1 = -2Q^3 + 2R^2 \equiv 6R \cdot \frac{R}{3} \pmod{Q}$$

and hence (i) holds with  $G = R/3$ . Then, using (31) again, we find

$$\begin{aligned} [F, F]_2 - (N+1)[F, G]_1 - (N+1)G^2 \\ = \frac{25}{9}(Q^3 - R^2) - \frac{10}{3}(Q^3 - R^2) - \frac{5}{9}R^2 = -\frac{5}{9}Q^3 \equiv 0 \pmod{Q^2} \end{aligned}$$

and hence (ii) also holds, proving that  $M_*$  is canonical with respect to the derivation  $\partial$  and element  $\Phi$  given by  $\partial(Q) = 0$ ,  $\partial(R) = -Q^2/2$ ,  $\Phi = -Q/36$ .

*Proof of the theorem.* The statement of the theorem indicates the proof. Assume first that  $M_*$  is canonical with respect to some  $\partial: M_* \rightarrow M_{*+2}$  and  $\Phi \in M_4$ , and choose any homogeneous element  $F \in M_N$ ,  $N > 0$ . Then properties (i) and (ii) in (b) hold with  $G = -\partial(F)$  because of the identities

$$\begin{aligned} [F, f]_1 - kfG &= [F, f]_{\partial, \Phi, 1} + kf\partial(F) = N\partial(f)F \quad (f \in M_k, k \geq 0), \\ [F, F]_2 + (N+1)[F, G]_1 - (N+1)G^2 &= (N(N+1)F\partial^2(F) - (N+1)^2\partial(F)^2 \\ &\quad + N^2(N+1)\Phi F^2) - (N+1)(N\partial(F)F - (N+1)F\partial^2(F)) - (N+1)(\partial(F))^2 \\ &= N^2(N+1)\Phi F^2. \end{aligned}$$

Conversely, suppose that  $M_*$  contains elements  $F \in M_N, G \in M_{N+2}$  for some  $N > 0$  satisfying (i) and (ii) (and with  $F$  not a zero-divisor), and define  $\partial$  and  $\Phi$  by (39). Then we claim that the brackets  $[\cdot]_{\partial, \Phi, *}$  induced by  $\partial$  and  $\Phi$  agree with the given bracket. As in earlier proofs, we can assume here that  $(M_*, [\cdot]_*)$  is a sub RC algebra of a standard RC algebra  $(R_*, [\cdot]_{D, *})$ , since the assertion to be proved is equivalent to a collection of universal identities for the brackets of RC algebras and such identities are true by definition if they are true for standard algebras. Now the larger algebra  $(R_*, [\cdot]_{D, *})$  is canonical, with derivation  $D$  and weight 4 element 0, so we have to show that in a ring with more than one choice of  $(F, G)$  as in (b) of the theorem, the induced bracket operations agree.

In fact, a little reflection shows that the key thing to check is that the property (b) in the theorem in a given RC algebra is independent of the choice of  $F$ , corresponding to the equivalence of (b) with the apparently much weaker (c). So now suppose that  $(F, G)$  satisfy (i) and (ii) and let  $\tilde{F} \in M_{\tilde{N}}$  be an arbitrary homogeneous element of  $M_*$ . We must show that there is an element  $\tilde{G} \in M_{\tilde{N}+2}$  so that  $(\tilde{F}, \tilde{G})$  also satisfy (ii). We may start by choosing any  $\tilde{G}$  which satisfies (ii), since we have already seen (in the "Remarks" above) that the truth or falsity of property (ii) is independent of the choice of  $\tilde{G}$  for a given  $\tilde{F}$ . We set

$$\tilde{G} = \frac{\tilde{N}G\tilde{F} - [F, \tilde{F}]_1}{NF}, \quad (40)$$

which belongs to  $M_{\tilde{N}+2}$  by property (i) of  $(F, G)$ . Then for  $f \in M_k$  we find

$$\partial_{F,G}(f) - \partial_{\tilde{F},\tilde{G}}(f) = \frac{\tilde{N}\tilde{F}[f, F]_1 + NF[\tilde{F}, f]_1 + k[\tilde{F}, F]_1}{N\tilde{N}F\tilde{F}} = 0$$

by the identity (10) of §2, and similarly  $\Phi_{F,G} - \Phi_{\tilde{F},\tilde{G}} = 0$  by virtue of the more complicated identity

$$\begin{aligned} N^2(N+1)F^2[\tilde{F}, \tilde{F}]_2 &= \tilde{N}^2(\tilde{N}+1)\tilde{F}_2[F, F]_2 - (N+1)(\tilde{N}+1)[F, \tilde{F}]_1^2 \\ &\quad + \tilde{N}(\tilde{N}+1)\tilde{F}[[F, \tilde{F}]_1, F]_1 \end{aligned} \quad (41)$$

which could have been (but was not) included in the list of universal identities in RC algebras given in §2. Thus the brackets constructed with  $\partial_{F,G}$  and  $\Phi_{F,G}$  are the same as those constructed from  $\tilde{F}$  and  $\tilde{G}$  chosen as in (40), and therefore the same as those constructed from any pair  $(\tilde{F}, \tilde{G})$  satisfying (i)–(ii) at all. (Changing  $G$  to  $G + \phi F$

changes  $\partial(f)$  ( $f \in M_k$ ) to  $\partial(f) + k\phi f$  and  $\Phi$  to  $\Phi + \phi^2 - \partial(\phi)$  but does not change the associated brackets, by the proof of Proposition 1.)  $\square$

We remark that the reason for the truth of the theorem is that we have the identities (10) and (41). The former says that, once we have fixed the multiplication (0th bracket) on an RC algebra, the first bracket for any two elements  $f, g \in M_*$  is determined once we have given the first brackets of  $f$  and  $g$  with a fixed homogeneous element  $F$  of  $M_*$  which is not a zero divisor. Similarly, the identity (41) tells us how to compute the second bracket  $[\tilde{F}, \tilde{F}]_2$  for any homogeneous  $\tilde{F} \in M_*$  (and hence also how to compute the second bracket  $[g, h]_2$  for any elements  $g, h \in M_*$ , by the usual polarization procedure for recovering a bilinear form from its associated quadratic form) knowing only the 0th and 1st brackets and the second bracket of  $F$  with itself. In other words, to specify the brackets in an RC algebra (assumed to contain one homogeneous non-zero divisor  $F$ ), we need to know only

1. the 0th bracket  $[f, g]_0$  for arbitrary  $f$  and  $g$ , which is arbitrary subject only to the conditions of bilinearity, associativity, and commutativity.
2. the 1st bracket of arbitrary elements  $f$  with the fixed element  $F$ , i.e. the derivation  $f \mapsto [f, F]_1$ , and
3. the 2nd bracket of  $F$  with itself, i.e. a single further element of  $M_*$ .

## 7. Other occurrences of Rankin–Cohen algebras

We end by raising the question where else RC algebras arise naturally in mathematics. One possible candidate, pointed out to me by T. Springer, is in invariant theory, where the algebras of invariants have natural bilinear operations called the “transvectant” or “Überschiebung” (cf [Sp], p. 66). These operations are indexed by integers  $n \geq 0$  and satisfy some universal identities of the general form of those occurring in §2, but they decrease rather than increasing the total weight (i.e., they send  $M_k \otimes M_l$  to  $M_{k+l-2n}$  rather than  $M_{k+l+2n}$ ). The relationship between the two types of algebraic structures remains to be determined. Another possibility are the so-called Moyal brackets in quantum theory, which are related to symplectic structures and seem to have similar algebraic properties to the brackets considered in this paper. Finally, the most natural source of interesting algebras with an infinite number of bilinear operations seems to be conformal field theories and more specifically vertex operator algebras. The axioms for vertex operator algebras as given in [Bo] or the appendix of [Ge] are different from ours, but discussions with Yu. Manin and W. Eholzer suggest that there may be a reformulation of the axioms of vertex operator algebras which is much closer to the RC algebras studied here. We hope to discuss this in a future publication.

## References

- [Bo] Borchers R, Vertex operator algebras, Kac-Moody algebras and the monster, *Proc. Natl. Acad. Sci. USA* **83** (1986) 3068–3071
- [Co] Cohen H, Sums involving the values at negative integers of L functions of quadratic characters, *Math. Ann.* **217** (1977) 81–94



- [EZ] Eichler M and Zagier D, The Theory of Jacobi Forms, *Prog. Math.* **55**, (Boston–Basel–Stuttgart, Birkhäuser) (1985)
- [Ge] Getzler E, Manin triples and  $N = 2$  superconformal field theory (preprint), MIT (1993)
- [IZ] Ibukiyama T and Zagier D, Higher spherical polynomials (in preparation)
- [Ku] Kuznetsov N V, A new class of identities for the Fourier coefficients of modular forms (in Russian), *Acta. Arith.* **27** (1975) 505–519
- [MZ] Manin Yu I and Zagier D, Automorphic pseudodifferential operators, (in preparation)
- [Ra] Rankin R A, The construction of automorphic forms from the derivatives of a given form, *J. Indian Math. Soc.* **20** (1956) 103–116
- [Sp] Springer T A, *Invariant Theory Lecture Notes* **585** (Berlin–Heidelberg–New York, Springer) (1977)
- [VZ] Rodriguez Villegas F and Zagier D, Square roots of central values of Hecke L-series, in “*Advances in Number Theory*, Proceedings of the third conference of the Canadian Number Theory Association” eds F Gouvea and N Yui (Oxford, Clarendon Press) (1993) 81–89

**Note added in proof.** Following a remark of W. Eholzer, it transpired that there is a further universal identity satisfied by the brackets in RC-algebras which is particularly simple and appealing: the multiplication on  $\oplus_k M_k$  defined by  $f * g = \sum_{n \geq 0} [f, g]_n$  is associative. That implies in turn infinitely many identities of the sort considered in §2 (possibly including all identities whose coefficients are independent of the weights, like (6), (8), (9), (11a) and (13)). Moreover, this multiplication turns out to be one of a whole one-parameter family of associative multiplications, all the rest of which do explicitly involve the weights of the forms involved, and one of which is the one arising from the correspondence with pseudodifferential operators discussed in §4. Details will be included in the paper [MZ].



# On Fourier coefficients of Maass cusp forms in 3-dimensional hyperbolic space

S RAGHAVAN and J SENGUPTA

School of Mathematics, Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Bombay 400 005, India

Dedicated to the memory of Professor K G Ramanathan

**Abstract.** In this article we establish the analogue of a theorem of Kuznetsov (theorem 6 of [3]) in the case of 3-dimensional hyperbolic space. We also consider a generalization of this result for higher dimensional hyperbolic spaces and discuss the relevant ingredients of a proof.

**Keywords.** Fourier coefficients, Maass cusp forms; 3-dimensional hyperbolic space; Kuznetsov theorem.

Let  $\mathbb{H}_3 = \{w = z + jy | z = x_1 + ix_2 \in \mathbb{C}, x_1, x_2, y \in \mathbb{R}, y > 0, ij = -ji, j^2 = -1 = i^2\}$  be the 3-dimensional hyperbolic space. The group  $PSL(2; \mathbb{C})$  acts on  $\mathbb{H}_3$  via the mappings  $w \rightarrow \gamma \langle w \rangle := (aw + b)(cw + d)^{-1}$  for  $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in PSL(2; \mathbb{C})$ . If  $R$  denotes the ring of integers in an imaginary quadratic field  $K$  over  $\mathbb{Q}$  of discriminant  $d = d_K < 0$  in  $\mathbb{Z}$ , then  $\Gamma := PSL(2; R)$  is a discontinuous group of homeomorphisms  $w \mapsto \gamma \langle w \rangle$  of  $\mathbb{H}_3$  onto  $\mathbb{H}_3$ , with a fundamental domain  $\mathcal{F}$ . On  $\mathbb{H}_3$ , we have the  $PSL(2; \mathbb{C})$ -invariant volume element  $dv = y^{-3} dx_1 dx_2 dy$ . Let  $\Gamma_\infty := \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \in \Gamma \right\}$  and  $\Gamma'_\infty := \left\{ \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \in \Gamma_\infty \right\}$ . For  $w \in \mathbb{H}_3$  and  $\gamma \in \Gamma$ , we write  $\gamma \langle w \rangle = z(\gamma \langle w \rangle) + jy(\gamma \langle w \rangle)$ .

For  $n$  in  $R^* := \{2a/\sqrt{d} | a \in R\}$ ,  $w \in \mathbb{H}_3$  and  $s$  in  $\mathbb{C}$  with  $\text{Re}(s) > 2$ , the Poincaré series  $P(w, s; n)$  is defined by

$$P(w, s; n) = \sum_{\gamma \in \Gamma'_\infty \backslash \Gamma} y(\gamma \langle w \rangle)^s \exp[-2\pi |n| y(\gamma \langle w \rangle) + 2\pi i \text{Re}(z(\gamma \langle w \rangle) \bar{n})]$$

For  $n = 0$ ,  $P(w, s; 0)$  is precisely the Eisenstein series denoted by  $E(w, s)$ . While  $P(w, s; n)$  is real analytic on  $\mathbb{H}_3$  and invariant under  $\Gamma$ , it is in  $L^2(\mathcal{F})$  only for  $n \neq 0$ ; on the other hand, it is an eigenfunction of the Laplacian  $\Delta$  on  $\mathbb{H}_3$  only for  $n = 0$ , with eigenvalue  $s(s - 2)$ .

The Eisenstein series  $E(w, s)$  can be continued meromorphically for all  $s$  in  $\mathbb{C}$ , the only singularity being a simple pole at  $s = 2$ . The Poincaré series  $P(w, s; n)$ , for  $n \neq 0$ , have a meromorphic continuation for all  $\text{Re}(s) > 1$  and are holomorphic there except possibly at a finite number of points  $s_i$ ; the finitely many  $\mu_i := s_i(2 - s_i)$  in  $(0, 1)$  are called the 'exceptional' eigenvalues (in the spectrum of  $\Delta$ ). The Eisenstein series  $E(w, 1 + it)$  for  $t$  in  $\mathbb{R}$  together with similar series for the cusps of  $\mathcal{F}$  different from  $\infty$  'span' the continuous spectrum of  $\Delta$ ; further, within  $[1, \infty)$ , there is also the discrete

spectrum of  $\Delta$  giving rise to the Maass waveforms  $u_l (l=1, 2, \dots)$  which are eigenfunctions for  $\Delta$  in  $L^2(\Gamma \backslash \mathbb{H}_3)$  with  $\Delta u_l + \lambda_l u_l = 0$  and  $\lambda_l \geq 1$ . For these  $u_l$  as well as the eigenfunctions corresponding to  $\mu_l$ , we have at the infinite cusp, the Fourier expansion

$$u_l(w) = \sum_{0 \neq m \in R^*} \rho_l(m) y K_{i\chi_l}(2\pi|m|y) \exp[2\pi i \operatorname{Re}(\bar{m}z)] \quad (1)$$

where  $K_\nu(\cdot)$  is the Macdonald function of order  $\nu$  and  $\chi_l := (u_l - 1)^{1/2}$  or  $(\lambda_l - 1)^{1/2}$ . If  $w'$  is the order of the group  $R^*$  of units of  $R$ , then the index  $[\Gamma_\infty : \Gamma'_\infty] = w'/2$ . In the case of fields  $K$  with class number 1, we know from ([1], Theorem 4.12) that the Eisenstein series  $E(w, s)$  has the Fourier expansion

$$E(w, s) = \frac{w'}{2} y^s + \frac{\pi w'}{(s-1)\sqrt{|d_K|}} \frac{\zeta_K(s-1)}{\zeta_K(s)} y^{2-s} + \frac{2\pi^s}{\sqrt{|d_K|} \Gamma(s) \zeta_K(s)} \times \sum_{0 \neq m \in R^*} |m|^{s-1} \sigma_1(-s) \left( \frac{m}{2} \sqrt{d_K} \right) y K_{s-1}(2\pi|m|y) \exp[2\pi i \operatorname{Re}(\bar{m}z)] \quad (2)$$

where  $\zeta_K$  is the Dedekind zeta function for  $K$  and  $\sigma_\nu(m) := \sum_{i|m} |i|^{2\nu}$ . We note that the expansion (2) is valid even for  $s = 1 + ir$ , for  $r$  in  $\mathbb{R}$ . Let us write  $\langle f, g \rangle := \int_{\mathcal{F}} f \bar{g} dv$  for  $f, g$  (measurable on  $\mathcal{F}$ ) and  $v(\mathcal{F}) := \int dv$ .

Our object is to obtain an estimate for the Fourier coefficients  $\rho_l(m)$  of  $u_l$ , in terms of  $m$  and  $\chi_l$ . This estimate can be deduced from an asymptotic formula for

$$A(X) := \sum_{|x_l| \leq X} \frac{|\rho_l(m)|^2}{1/|\Gamma(1 - i\chi_l)|^2} \quad (3)$$

More specifically, we shall prove the following.

**Theorem.** For  $A(X; m) = A(X)$  defined by (3), we have, as  $X \rightarrow \infty$ ,

$$A(X; m) = \frac{y(\mathcal{F}_K)}{\pi^3 |d_K| c_1} X^3 + O(X^{5/2} |m|^{1+\varepsilon}) \quad (4)$$

for any  $\varepsilon > 0$  with the  $O$ -constant depending on  $\varepsilon$ , where  $c_1 := 2 \int_0^\infty \frac{u}{\sinh \pi u} du$ . As a consequence, we obtain

$$\rho_l(m) = \mathcal{O}(\chi_l^{3/4} e^{\pi \chi_l/2} |m|^{1/2+\varepsilon}) \quad (5)$$

and  $O$ -constant depending at most on  $\varepsilon$ .

**Remark.** Surprisingly, the contribution to  $A(X)$  from the exceptional eigenvalues is very much under control here (unlike in the problem of sums of Kloosterman sums).

The proof of the theorem follows Kuznetsov [3] and at the same time, hopefully helps in clearing a few obscurities. It rests, of course, on an identity arising from the computation of the inner product

$$\langle P(\cdot, 2 + it; m), P(\cdot, 2 + it; n) \rangle \text{ for } t \in \mathbb{R} \text{ and } m, n \neq 0 \text{ in } R^* \quad (6)$$

in two different ways, namely via the Parseval relation for  $L^2(\mathcal{F})$  and also using the

$$\langle P(., s; m), E(., 1 + ir) \rangle =$$

$$\frac{2^{1+2ir} \pi^{3/2}}{\sqrt{|d_K|} (4\pi m)^{s-1+ir}} \frac{\Gamma(s-1+ir) \Gamma(s-1-ir) \sigma_{ir} \left( \frac{m}{2} \sqrt{d_K} \right)}{\Gamma(1-ir) \Gamma(s-\frac{1}{2}) \zeta_K(1-ir)} \quad (7)$$

The left hand side is precisely

$$\begin{aligned} & \int_{\mathcal{F}} \sum_{y \in \Gamma_{\infty} \backslash \Gamma} (y(\gamma \langle w \rangle))^s \exp[-2\pi|m|y(\gamma \langle w \rangle) + 2\pi i \operatorname{Re}(z(\gamma \langle w \rangle) \bar{m})] \overline{E(w, 1 + ir)} dv \\ &= \int_{\Gamma_{\infty}' \backslash \mathbb{H}_3} y^s \exp[-2\pi|m|y] \overline{E(w, 1 + ir)} \exp[2\pi i \operatorname{Re}(z \bar{m})] dv \\ &= \int_0^{\infty} y^{s-3} \exp[-2\pi|m|y] dy \int_{R \setminus \mathbb{C}} \overline{E(w, 1 + ir)} \exp[-2\pi i \operatorname{Re}(z \bar{m})] dx_1 dx_2 \\ &= \int_0^{\infty} y^{s-3} \exp[-2\pi|m|y] \frac{2\pi^{1-ir} |m|^{-ir} \sigma_{ir} \left( \frac{m}{2} \sqrt{d_K} \right)}{\sqrt{|d_K|} \Gamma(1-ir) \zeta_K(1-ir)} y K_{ir}(2\pi|m|y) dy, \text{ by (2)} \\ &= \frac{2\sqrt{\pi} \Gamma(s-1+ir) \Gamma(s-1-ir)}{|d_K|^{1/2} (4\pi|m|)^{s-1} \Gamma(s-1/2)} \frac{\pi^{1-ir}}{|m|^{ir}} \frac{\sigma_{ir} \left( \frac{m}{2} \sqrt{d_K} \right)}{\Gamma(1-ir) \zeta_K(1-ir)} \end{aligned}$$

which proves (7). Using now (7) with  $m, n \neq 0$  in  $R^*$ ,

$$\begin{aligned} & \int_{-\infty}^{\infty} \langle P(., s; m), E(., 1 + ir) \rangle \overline{\langle P(., s; n), E(., 1 + ir) \rangle} dr \\ &= \frac{4\pi^3 (4\pi)^{2-s-\bar{s}}}{|m|^{s-1} |n|^{\bar{s}-1} |d_K|} \\ & \quad \times \int_{-\infty}^{\infty} \left| \frac{n}{m} \right|^{ir} \frac{|\Gamma(s-1+ir)|^2 |\Gamma(s-1-ir)|^2 \sigma_{ir} \left( \frac{m}{2} \sqrt{d_K} \right) \overline{\sigma_{ir} \left( \frac{n}{2} \sqrt{d_K} \right)}}{|\Gamma(1-ir)|^2 |\Gamma(s-1/2)|^2 |\zeta_K(1+ir)|^2} dr. \end{aligned} \quad (8)$$

On the other hand, we know from Sarnak [4], that, for  $t$  in  $\mathbb{R}$ ,

$$\begin{aligned} & \langle P(., 2 + it; m), u_t(\cdot) \rangle \overline{\langle P(., 2 + it; n), u_t(\cdot) \rangle} \\ &= \frac{|d_K| \rho_t(m) \overline{\rho_t(n)}}{16\pi |m|^{1+it} |n|^{1-it}} \frac{|\Gamma(1+i(t+\chi_t))|^2 |\Gamma(1+i(t-\chi_t))|^2}{|\Gamma(3/2+it)|^2}. \end{aligned} \quad (9)$$

Using (8) and (9), the Parseval relation gives us immediately that the inner product in (6) is nothing but

$$\begin{aligned} & \frac{|d_K|}{16\pi|m|^{1+it}|n|^{1-it}} \sum_{l=1}^{\infty} \rho_l(m) \overline{\rho_l(n)} \frac{|\Gamma(1+i(t+\chi_l))|^2 \Gamma(1+i(t-\chi_l))|^2}{|\Gamma(3/2+it)|^2} \\ & + \frac{4\pi^3 |\Gamma(3/2+it)|^{-2}}{(4\pi)^2 |d_K| |m|^{1+it} |n|^{1-it}} \times \int_{-\infty}^{\infty} \left| \frac{n}{m} \right|^{ir} \\ & \times \frac{|\Gamma(1+i(t+r))|^2 |\Gamma(1+i(t-r))|^2 \sigma_{ir}\left(\frac{m}{2}\sqrt{d_K}\right) \overline{\sigma_{ir}\left(\frac{n}{2}\sqrt{d_K}\right)}}{|\Gamma(1-ir)|^2 |\zeta_K(1-ir)|^2} \frac{dr}{4\pi} \quad (10) \end{aligned}$$

Using the Fourier expansion of the Poincaré series  $P(., 2+it, .)$  from [4], we see that the inner product in (6) also equals

$$\begin{aligned} & \frac{\delta_{m,n} v(\mathcal{F}_R)}{4\pi^2 (|m|+|n|)^2} + \sum_{\substack{0 \neq c \in R \\ c \bmod \pm 1}} \frac{S(m, n; c)}{|c|^{4+2it}} \int_0^{\infty} \frac{dy}{y^{1+2it}} \int_{\mathbb{R}^2} \\ & \times \exp \left[ -2\pi \left( |n|y + \frac{|m|}{y|c|^2(1+|z|^2)} \right) + 2\pi i \operatorname{Re} \left( -\frac{z\bar{m}}{c^2 y(1+|z|^2)} - y\bar{n}z \right) \right. \\ & \left. \times \frac{dx_1 dx_2}{(1+|z|^2)^{2+it}} \right] \quad (11) \end{aligned}$$

where

$$S(m, n; c) = \sum_{\substack{a \bmod c \\ a\bar{a} \equiv 1 \pmod{c}}} \exp[2\pi i \operatorname{Re}((\bar{m}a + \bar{n}\bar{a})/c)].$$

For the Kloosterman sums  $S(m, m; c)$  we have from [2], the estimate similar to Weil's, viz. for every  $\varepsilon > 0$ ,

$$|S(m, m; c)| \leq |c|^{1+\varepsilon} |(m\sqrt{d_K}, c)| \sigma_0(c/(\bar{m}\sqrt{d_K}, c)) \quad (12)$$

On the same lines as in Kuznetsov [3], let us introduce

$$H(r, t) = \frac{|\Gamma(1+i(t+r))|^2 |\Gamma(1+i(t-r))|^2}{\pi |\Gamma(1-ir)|^2}$$

and

$$h_Y(r) = \int_0^Y \cosh(\pi t) H(r, t) dt, \text{ for } Y > 0 \text{ and } r \text{ in } \mathbb{C} \text{ with } |\operatorname{Im} r| < 1.$$

Moreover, let us multiply the two equal expressions (10) and (11) for the inner product in (6) with  $m = n$ , throughout by

$$\frac{16\pi^2 |m|^2}{|d_K|} \left( \frac{1}{4} + t^2 \right) = \frac{16\pi^2 |m|^2}{|d_K|} \frac{\cosh(\pi t) |\Gamma(3/2+it)|^2}{\pi}$$

and then integrate both the resulting expressions with respect to  $t$  from 0 to  $Y$ . IV

then obtain

$$\begin{aligned}
& \sum_l \frac{|\rho_l(m)|^2}{1/|\Gamma(1-i\chi_l)|^2} h_Y(\chi_l) + \frac{\pi}{|d_K|^2} \int_0^\infty \cosh(\pi t) dt \int_{-\infty}^\infty \frac{H(r, L) \left| \sigma_{ir} \left( \frac{m}{2} \sqrt{d_K} \right) \right|^2}{|\zeta_K(1+ir)|^2} dr = \\
& = \frac{v(\mathcal{F}_R)}{|d_K| \pi} \left( \frac{1}{3} Y^3 + \frac{1}{4} Y \right) + \frac{16\pi|m|^2}{|d_K|} \sum_{\substack{0 \neq c \in R \\ c \bmod \pm 1}} |c|^{-1} S(m, m; c) \times \\
& \times \int_0^Y \int_0^\infty \int_{\mathbb{R}^2} \left( \frac{1}{4} + t^2 \right) (y^2 |\delta_1|)^{-it} \\
& \exp[-2\pi|m|(y + 1/(y|\delta_1|)) - 2\pi i \operatorname{Re}(\bar{m}z(y + 1/(y\delta_1)))] dt \frac{dy}{y} \frac{dx_1 dx_2}{(1+x_1^2+x_2^2)^2}
\end{aligned} \tag{13}$$

using the abbreviation  $\delta_1$  for  $c^2(1+x_1^2+x_2^2)$  and recalling that  $z = x_1 + ix_2$ .

To derive an asymptotic formula for  $A(X)$  as  $X$  tends to infinity, we first need to estimate  $h_Y(r)$  before employing (13). We see actually that

$$h_Y(r) = \int_0^Y \frac{t^2 - r^2}{2r} \left( \frac{1}{\sinh(\pi t - \pi r)} - \frac{1}{\sinh(\pi t + \pi r)} \right) dt$$

and then it is not difficult to find that indeed

$$h_\infty(r) = 2 \int_0^\infty \frac{u}{\sinh(\pi u)} du$$

(which positive constant we denote by  $c_1$ ). Again, as in [2], we have for  $1 \leq r \leq Y - \log Y$ ,

$$c_1 > h_Y(r) = c_1 + O(Y^{-\pi/2} + \exp(-\pi r)) \tag{14}$$

Further, for  $r \geq Y + \log Y$ ,

$$h_Y(r) \ll (r - Y) \exp(-\pi(r - Y)). \tag{15}$$

We next estimate the series

$$\frac{16\pi|m|^2}{|d_K|} \sum_c |c|^{-4} S(m, m; c) I(c)$$

on the right hand side of (13), denoting the relevant integral therein by  $I(c)$ . This integral may be rewritten as

$$\begin{aligned}
I(c) &= \int_0^\infty \frac{dy}{y} \int_0^\infty \frac{r dr}{(1+r^2)^2} \int_0^{2\pi} \\
&\times \exp[-2\pi|m|(\delta(y + y^{-1}) - 2\pi i \operatorname{Re}(re^{i\theta} \bar{m}\delta(y + y^{-1} \exp(-2i \arg c))))] d\theta \times \\
&\times \int_0^Y \left( \frac{1}{4} + t^2 \right) \exp[-2it \log y] dt.
\end{aligned} \tag{16}$$

(after replacing  $y/\delta$  by  $y$ , with  $\delta := 1/|\delta_1|^{1/2} = 1/(|c|(1+r^2)^{1/2})$ ). The integral with respect to  $t$  occurring in (16) is trivially  $O(Y^3)$ ; however, for  $y \neq 1$ , it can be seen to be  $O(Y^2/|\log y|)$ , on using integration by parts. The integration with respect to  $y$  can be broken up into integration over  $A_1 := (0, 1/w_Y)$ ,  $A_2 := [1/w_Y, w_Y]$  and  $A_3 := (w_Y, \infty)$  writing  $w_Y$  for  $\exp(1/Y^\varepsilon)$ , for any fixed  $\varepsilon > 0$ . We are then led to the estimation

$$\begin{aligned} I(c) &\ll Y^2 \int_0^\infty \frac{r dr}{(1+r^2)^2} \left( \int_{A_1} + \int_{A_3} \right) \exp[-2\pi|m|\delta(y+y^{-1})] \frac{dy}{y|\log y|} + \\ &\quad + Y^3 \int_0^\infty \frac{r dr}{(1+r^2)^2} \int_{A_2} \exp[-2\pi|m|\delta(y+y^{-1})] \frac{dy}{y} \\ &\ll Y^2 \int_0^\infty \frac{r}{(1+r^2)^2} Y^\varepsilon \left( \frac{|c|(1+r^2)^{1/2}}{|m|} \right)^{1-\beta} \\ &\quad \times dr + Y^3 \int_0^\infty \frac{r}{(1+r^2)^2} Y^{-\varepsilon} \left( \frac{|c|(1+r^2)^{1/2}}{|m|} \right)^{1-\beta} dr, \end{aligned}$$

for any  $\beta$  in  $(0, 1)$ . Choosing now  $\varepsilon = 1/2$ , we obtain that

$$I(c) \ll \frac{Y^{5/2}}{(|m|/|c|)^{1-\beta}} \int_0^\infty \frac{r dr}{(1+r^2)^{(3+\beta)/2}} \ll Y^{5/2} (|c|/|m|)^{1-\beta}$$

for any  $\beta$  in  $(0, 1)$ . This together with the estimate (12) implies, for the second term on the right hand side of (13), the bound, for any  $\beta$  in  $(0, 1)$ :

$$\frac{16\pi|m|^2}{|d_K|} \sum_{\substack{0 \neq c \in R \\ c \bmod \pm 1}} |c|^{-4} S(m, m; c) I(c) \ll Y^{5/2} \sigma_{(1+\beta)/2}(m\sqrt{d_K}). \quad (17)$$

For the contribution to the first term on the left hand side of (13) from all the exceptional  $\chi_l := (\mu_l - 1)^{1/2}$ , we derive the following estimate, namely

$$\sum_{\chi_l = (\mu_l - 1)^{1/2}} |\rho_l(m)|^2 h_Y(\chi_l) |\Gamma(1 - i\chi_l)|^2 \ll |m|^{1+\varepsilon} \text{ for any } \varepsilon > 0. \quad (18)$$

For this purpose, we note that these  $\chi_l$  are all purely imaginary and further know from Sarnak ([4], Theorem 3.1) that  $|\chi_l| \leq 1/2$ ; hence  $H(\chi_l, t)$  and  $h_Y(\chi_l)$  are well-defined. The asymptotic formula  $\Gamma(x + iy) \sim \sqrt{2\pi} \exp[-\pi|y|/2] |y|^{x-1/2}$  for  $|y| \rightarrow \infty$  implies that  $H(\chi_l, t) \sim 2\pi|t|^2 \exp[-2\pi|t|]/|\Gamma(1 - i\chi_l)|^2$  as  $t \rightarrow \infty$ . Consequently, for every  $Y > 0$ , we have

$$h_Y(\chi_l) / \{1/|\Gamma(1 - i\chi_l)|^2\} \ll \int_0^Y t^2 \exp[-2\pi t] \cosh(\pi t) dt < 4/\pi^3.$$

Now, for  $Y \geq 1$ ,

$$\min_{\chi_l = (\mu_l - 1)^{1/2}} \frac{h_Y(\chi_l)}{1/|\Gamma(1 - i\chi_l)|^2} \geq \min_{\chi_l = (\mu_l - 1)^{1/2}} \frac{h_1(\chi_l)}{1/|\Gamma(1 - i\chi_l)|^2} =: C(> 0)$$



$$C \sum_{\chi_l = (\mu_l - 1)^{1/2}} |\rho_l(m)|^2 \leq \sum_l \frac{|\rho_l(m)|^2 h_1(\chi_l)}{1/|\Gamma(1 - i\chi_l)|^2} \ll |m|^{1+\varepsilon}, \quad \text{by (13) and (17).}$$

This proves (18).

For the proof of our theorem, we shall, in the light of (18), totally ignore the presence of the  $\mu_l$ 's, in the sequel.

Setting  $Y = X + \log X$  in (13), the left hand side of (13) is  $\geq (c_1 - c_2 X^{-\pi/2})A(X) - c_3 \sum_{|\chi_l| \leq X} |\rho_l(m)|^2 |\exp[-\pi\chi_l]|\Gamma(1 - i\chi_l)|^2$  for certain positive constants  $c_2$  and  $c_3$ , on using (14) as well as the positivity of the terms on the left hand side of (13). Hence, for any  $\varepsilon > 0$ , we obtain, with a constant  $c_4 > 0$ ,

$$\frac{\pi 3|d_K|}{v(\mathcal{F}_R)} c_1 A(X) \leq X^3 + c_4 X^{5/2} |m|^{1+\varepsilon}. \quad (19)$$

We next substitute  $Y = X - \log X$  in (13) and noting that  $Y + \log Y \approx X$  and  $Y - \log Y \geq X - 2\log X$ , we get by (14) that

$$\begin{aligned} & \sum_{\chi_l \leq X - 2\log X} |\rho_l(m)|^2 \frac{(c_1 + O((X - \log X)^{-\pi/2} + \exp(-\pi\chi_l)))}{1/|\Gamma(1 - i\chi_l)|^2} + \\ & + \sum_{X - 2\log X < \chi_l \leq X} |\rho_l(m)|^2 h_Y(\chi_l) + \\ & + O\left(\sum_{\chi_l > X} \exp(-\pi(\chi_l - X + \log X)) \frac{|\rho_l(m)|^2 (\chi_l - X + \log X)}{1/|\Gamma(1 - i\chi_l)|^2}\right) + I_1 = \\ & = \frac{v(\mathcal{F}_R)}{\pi|d_K|} \left\{ \frac{(X - \log X)^3}{3} + \frac{X - \log X}{4} \right\} + O((X - \log X)^{5/2} |m|^{1+\varepsilon}) \end{aligned}$$

where  $I_1$  denotes the second term on the left hand side of (13). From (14), (15) and the estimate  $|\zeta_K(1 + ir)|^{-1} = O(|r|^\varepsilon)$  for every  $\varepsilon > 0$  as  $|r| \rightarrow \infty$ , we see that  $I_1 = O(X^{1+\varepsilon} |m|^\varepsilon)$ , while the right hand side of the equality above is  $\geq \frac{v(\mathcal{F}_R)}{\pi 3|d_K|} X^3 - c_5 X^{5/2} |m|^{1+\varepsilon}$  for a positive constant  $c_5$ . But now, by (13) and (17),

$$\begin{aligned} & \sum_{X - 2\log X \leq \chi_l \leq X} |\rho_l(m)|^2 \frac{h_Y(\chi_l)}{1/|\Gamma(1 - i\chi_l)|^2} = O(X^3 - (X - 2\log X)^3) + \\ & + O(X^{5/2} |m|^{1+\varepsilon}) \\ & = O(X^{5/2} |m|^{1+\varepsilon}) \end{aligned}$$

Therefore

$$c_1 A(X) \geq c_1 A(X - 2\log X) \geq \frac{v(\mathcal{F}_R)}{3|d_K|\pi} X^3 - c_6 X^{5/2} |m|^{1+\varepsilon},$$

for a positive constant  $c_6$ . This together with (19) proves the Theorem. The estimate (5) follows on applying the usual difference argument to (4).

*Remarks (i)* In principle, it should be possible to derive the estimate (5) through the general theory of automorphic forms on  $GL(2)$  but our proof is elementary.

(ii) An estimate similar to (5) with  $|m|^{1+\varepsilon}$  in place of  $|m|^{1/2+\varepsilon}$  may be obtained in the case of Maass cusp forms on 4-dimensional hyperbolic space (the order of integral Hurwitz quaternions replacing the ring  $R$ ).

## Appendix

In this appendix we indicate how one can proceed to generalise the result of the theorem above to higher-dimensional hyperbolic spaces. We will utilise the exposition in [5] for our basic set up in this case.

Let  $q$  be a non-degenerate quadratic form on a  $k$ -dimensional vector space  $E$  over  $\mathbf{Q}$  and  $\mathcal{C}(q)$ , the associated Clifford algebra (see [5]), identifying  $\mathbf{Q}$  and  $E$  with their canonical images in  $\mathcal{C}(q)$ ; for  $k=0$ ,  $\mathcal{C}(q) = \mathbf{Q}$ . Taking an orthogonal basis  $\{e_1, \dots, e_k\}$  for  $E$  over  $\mathbf{Q}$  with respect to  $q$ , we have

$$e_p^2 = q(e_p)(p = 1, 2, \dots, k), \quad e_l e_m = -e_m e_l (1 \leq l \neq m \leq k).$$

For any subset  $(M = \{e_{v_1}, \dots, e_{v_r} | v_1 < v_2 < \dots < v_r\})$  of the basis  $\{e_1, \dots, e_k\}$ , define  $e_M := e_{v_1} \cdots e_{v_r}$  and  $e_\emptyset := 1$  for the empty set  $\emptyset$ . Then these  $e_M$  form a basis for  $\mathcal{C}(q)$  over  $\mathbf{Q}$ . We have three  $\mathbf{Q}$ -linear involutions  $x \mapsto x'$ ,  $x \mapsto \bar{x}$  and  $x \mapsto x^*$  on  $\mathcal{C}(q)$  reducing to the identity map on  $\mathbf{Q}$  such that for any  $e_M$  as above  $e'_M = (-1)^r e_M$ ,  $\bar{e}_M = (-1)^{(r^2+r)/2} e_M$  and  $e_M^* = (-1)^{(r^2-r)/2} e_M$ . Further for any  $x, y$  in  $\mathcal{C}(q)$ ,

$$\bar{x}\bar{y} = \bar{y}\bar{x} \text{ and } (xy)^* = y^*x^*.$$

We have a *trace* map  $\text{tr}: \mathcal{C}(q) \rightarrow \mathcal{C}(q)$  defined by  $\text{tr}(x) = x + \bar{x}$ . When  $q = -I_k$ , the negative of the unit quadratic form  $I_k$  on  $E$ , we have on  $V_{k+1} := \mathbf{Q}.1 \oplus E \subset \mathcal{C}(q)$ , a scalar product  $\langle v, w \rangle := \frac{1}{2} \text{tr}(v\bar{w})$  for all  $v, w$  in  $V_{k+1}$ , so that  $\{1, e_1, \dots, e_k\}$  becomes an orthonormal basis for  $V_{k+1}$ . For any  $x = \sum_M \lambda_M e_M$  in  $\mathcal{C}(q)$  with  $\lambda_M \in \mathbf{Q}$  or more generally

for  $x = \sum_M \lambda_M e_M = \sum_M \lambda_M e_M \otimes 1$  in  $\mathcal{C}(q) \otimes \mathbf{R}$  with  $\lambda_M$  in  $\mathbf{R}$ , we know that  $|x| :=$

$\left( \sum_M \lambda_M^2 \right)^{1/2}$  defines the euclidean norm of  $x$ . Further we have  $|v|^2 = v\bar{v} = \bar{v}v$  whenever  $v \neq 0$  in  $\mathcal{C}(q) \otimes \mathbf{R}$  has the property that there exists a  $\mathbf{Q}$ -linear automorphism  $\varphi_v: V_{k+1} \rightarrow V_{k+1}$  such that  $v x = \varphi_v(x) v'$  for all  $x$  in  $V_{k+1}$ . We denote the algebra  $\mathcal{C}(q) \otimes \mathbf{R}$  and the vector space  $V_{k+1} \otimes \mathbf{R}$  resulting by base change respectively from  $\mathcal{C}(q)$  and  $V_{k+1}$ , also by  $\mathcal{C}(q)$  and  $V_{k+1}$  again. Given a lattice  $L$  in  $V_{k+1}$ , the dual lattice  $L^\#$  is defined by  $L^\# := \{y \in V_{k+1} | \langle x, y \rangle \in \mathbf{Z} \text{ for every } x \text{ in } L\}$ .

Let  $\mathbf{H}^{k+2}$  be the  $(k+2)$ -dimensional hyperbolic space given by the upper half-space

$$\{w = x_0 + x_1 e_1 + x_2 e_2 + \dots + x_k e_k + x_{k+1} e_{k+1} | \\ x_0, x_1, x_2, \dots, x_{k+1} \in \mathbf{R}, x_{k+1} > 0\}.$$

We write  $w = Z + r e_{k+1}$  or more precisely, for  $P = (x_0, x_1, \dots, x_{k+1})$  corresponding to  $w = x_0 + x_1 e_1 + \dots + x_k e_k + x_{k+1} e_{k+1}$  in  $\mathbf{H}^{k+2}$ ,  $Z = Z(P) := x_0 + x_1 e_1 + \dots + x_k e_k$  and  $r = r(P) = x_{k+1} > 0$ . Then  $|W(P)|^2 := |Z(P)|^2 + (r(P))^2$  or simply  $|w|^2 := |Z|^2 + r^2$ .

We have on  $\mathbf{R}^{k+2}$ , a Riemannian metric  $ds^2 = dx_{k+1}^2(dx_0^2 + \dots + dx_{k+1}^2)$  and associated volume element  $dv := x_{k+1}^{-(k+2)} dx_0 \wedge dx_1 \wedge \dots \wedge dx_{k+1}$ ; the corresponding Laplace-Beltrami operator  $\Delta$  is given by

$$\Delta := x_{k+1}^2 \left( \frac{\partial^2}{\partial x_0^2} + \dots + \frac{\partial^2}{\partial x_{k+1}^2} \right) - k x_{k+1} \frac{\partial}{\partial x_{k+1}},$$

For the Clifford algebra  $\mathcal{C}(q)$  over the base field  $K = \mathbf{R}$  or  $\mathbf{Q}$  and  $q = -I_K$ , the *Vahlen group*  $SV_k(K)$  is defined by

$$SV_k(K) := \left\{ \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \mid \begin{array}{l} \text{(i)} \quad \alpha, \beta, \gamma, \delta \in \mathcal{C}(q) \text{ with } \alpha\delta^* - \beta\gamma^* = 1 \\ \text{(ii)} \quad \alpha\beta^* = \beta\alpha^*, \gamma\delta^* = \delta\gamma^* \\ \text{(iii)} \quad \alpha\bar{\alpha}, \beta\bar{\beta}, \gamma\bar{\gamma}, \delta\bar{\delta} \in K \\ \text{(iv)} \quad \alpha\bar{\gamma}, \beta\bar{\delta} \in V_{k+1} \\ \text{(v)} \quad \alpha x\bar{\beta} + \beta\bar{x}\bar{\alpha}, \gamma x\bar{\delta} + \delta\bar{x}\bar{\gamma} \in K, \forall x \in V_{k+1} \\ \text{(vi)} \quad \alpha x\bar{\delta} + \beta\bar{x}\bar{\gamma} \in V_{k+1}, \forall x \in V_{k+1} \end{array} \right\}$$

For  $K = \mathbf{R}$ ,  $SV_0 = SL_2(\mathbf{R})$  and  $SV_1 = SL_2(\mathbf{C})$ . Let  $J$  be a  $\mathbf{Z}$ -order in the  $\mathbf{Q}$ -algebra  $\mathcal{C}(q)$  (i.e. a subring containing 1 and a  $\mathbf{Q}$ -basis of  $\mathcal{C}(q)$ , with the underlying additive group finitely generated) which is stable under the involutions  $*$  and  $'$  of  $\mathcal{C}(q)$ . By  $\Gamma := SV_k(J)$ , we mean the subgroup  $\left\{ \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in SV_k(\mathbf{Q}) \mid \alpha, \beta, \gamma, \delta \in J \right\}$ . The Vahlen group  $SV_k(\mathbf{R})$  acts on  $\mathbf{H}^{k+2}$  as orientation preserving isometries through the maps

$$P \mapsto \sigma P := (\alpha P + \beta)(\gamma P + \delta)^{-1} \text{ for } \sigma = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in SV_k(\mathbf{R}).$$

For any  $w = w(P) = Z(P) + r(P)e_{k+1} \in \mathbf{H}^{k+2}$ , we have correspondingly for  $w(\sigma P) = Z(\sigma P) + r(\sigma P)e_{k+1}$ ,

$$\begin{aligned} Z(\sigma P) &= \frac{(\alpha Z + \beta)(\overline{\gamma Z + \delta}) + \alpha\bar{\gamma}r^2}{|\gamma Z + \delta|^2 + |\gamma|^2 r^2}, \quad \text{with } Z := Z(P), \quad r := r(P) \\ &= \alpha\gamma^{-1} - (\gamma^*)^{-1} \frac{(\overline{\gamma Z + \delta})}{|\gamma Z + \delta|^2 + |\gamma|^2 r^2} \quad \text{whenever } \gamma \neq 0, \end{aligned}$$

and  $r(\sigma P) = r(P)/(|\gamma Z + \delta|^2 + |\gamma|^2 r^2)$ . By  $\Gamma'_\infty$ , we mean the subgroup  $\left\{ \begin{pmatrix} 1 & w \\ 0 & 1 \end{pmatrix} \in \Gamma \right\}$ .

We also assume, for simplicity, that  $\Gamma$  has only one cusp viz.  $\infty$ , for the above action on  $\mathbf{H}^{k+2}$ .

If  $\Lambda := J \cap V_{k+1}$  for a given  $\mathbf{Z}$ -order  $J$  in  $\mathcal{C}(q)$  over  $\mathbf{Q}$ , then  $\Lambda$  is a lattice in  $V_{k+1}$ . Then, for any  $\mu$  in the dual lattice  $\Lambda^\#$  and  $s$  in  $\mathbf{C}$  with  $\text{Re}(s) > k+1$ , the Poincaré series  $U_\mu(., s)$  is defined by

$$U_\mu(P, s) := \sum_{\sigma \in \Gamma'_\infty \backslash \Gamma} r(\sigma P)^s e(i|\mu|r(\sigma P) + \langle Z(\sigma P), \mu \rangle) \quad (1)$$

where  $e(\theta) := \exp(2\pi i \theta)$  for  $\theta \in \mathbf{C}$  and  $i = \sqrt{-1} \in \mathbf{C}$  (with  $\arg i = \pi/2$ ). (see [5]). This series

converges absolutely to compact subsets of  $\mathbf{H}^{k+2} \times \{s \in \mathbb{C} \mid \operatorname{Re}(s) > k+1\}$  and  $U_0(P, s)$  is actually an Eisenstein series. If  $\mu \neq 0$  in  $\Lambda^\#$  and  $\operatorname{Re}(s) > k+1$ ,  $U_\mu(\cdot, s) \in L^2(\Gamma \backslash \mathbf{H}^{k+2})$ . For  $\operatorname{Re}(s) > k+1$  again,  $U_\mu(\cdot, s)$  satisfies the differential equation

$$(-\Delta - s(k+1-s))U_\mu(\cdot, s) = 2\pi|\mu|(2s-k)U_\mu(\cdot, s+1) \quad (2)$$

which implies immediately that  $U_\mu(\cdot, s)$  has a meromorphic continuation to the domain given by  $\operatorname{Re}(s) > k$ ; further, it has no pole at  $s = k+1$  and indeed, from ([5], Theorem 10.1) it even follows that  $U_\mu(\cdot, s)$  is holomorphic in  $s$  for  $\operatorname{Re}(s) > k+1/2$ . The possible poles of  $U_\mu(\cdot, s)$  in  $((k+1)/2, k+1)$  correspond to the values of  $s > (k+1)/2$  for which  $s(k+1-s)$  is an ("exceptional") eigenvalue of  $-\Delta$ .

The proof of the proposed generalization rests as before, on an identity arising from the computation of the inner product

$$\langle U_\mu(\cdot, k+1+it), U_\mu(\cdot, k+1+it) \rangle \quad \text{for } \mu \neq 0 \text{ in } \Lambda^\#, \quad t \in \mathbb{R}$$

in two different ways, namely through the Parseval relation in  $L^2(\Gamma \backslash \mathbf{H}^{k+2})$  or by using the Fourier expansion of the given Poincaré series as described below.

Let  $C_{\mu, v}$ , for given  $\mu, v \in \Lambda^\#$ , denote the number of  $\alpha$  in  $J$  such that  $\alpha\bar{\alpha} = 1$  and  $\varphi_\alpha^*(\mu) = v$  where  $\varphi_\alpha^*$  is the map dual (with respect to  $\langle, \rangle$ ) to the map  $\varphi_\alpha$  defined by  $\alpha x = \varphi_\alpha(x)\alpha'$  for every  $x$  in  $V_{k+1} \otimes \mathbb{R}$ . We note that  $C_{\mu, v}$  is  $O(1)$  for all  $\mu, v$ . For all  $s$  in  $\mathbb{C}$  with  $\operatorname{Re} s > k+1$ , we have from [5] the following Fourier expansion

$$\begin{aligned} v(\mathcal{F}_\Lambda)U_\mu(w, s) = & \sum_{v \in \Lambda^\#} e(\langle Z, v \rangle) \left\{ C_{\mu, v} r^s \exp(-2\pi|\mu|r) v(\mathcal{F}_\Lambda) \right. \\ & + r^{k+1-s} \sum_{\gamma \bar{\gamma} \neq 0} \frac{S(\mu, v; \gamma)}{(\gamma \bar{\gamma})^s} \int_{V_{k+1}} \frac{\exp\left[\frac{-2\pi|\mu|}{r|\gamma|^2(1+|Z|^2)}\right]}{(1+|Z|^2)^s} e\left(-\langle \frac{(\gamma^*)^{-1}\bar{Z}\gamma^{-1}}{r(1+|Z|^2)}, \mu \rangle \right. \\ & \left. \left. - r\langle Z, v \rangle \right) dZ \right\} \quad (3)_A \end{aligned}$$

where  $v(\mathcal{F}_\Lambda)$  is the volume of fundamental domain  $\mathcal{F}_\Lambda = V_{k+1}/\Lambda$ ,  $\delta_{..}$  denotes the Kronecker delta and the generalized Kloosterman sum  $S(\mu, v, \gamma)$  for  $\gamma \neq 0$  in  $J$  is defined by

$$S(\mu, v; \gamma) := \sum_{(x, y) \in D(\gamma)} e(\langle xy^{-1}, \mu \rangle + \langle \gamma^{-1}y, v \rangle)$$

with  $D(\gamma) := \{(\alpha, \delta) \mid \sigma = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \text{ for fixed } \gamma \text{ lie in distinct double cosets of } \Gamma \text{ modulo } \Gamma_\infty'\}$ . For  $(\alpha, \delta) \in D(\gamma)$ , we have  $\alpha\delta^* - 1 \in J\gamma^*$  by definition.

Let  $G(r, v)$  denote the series occurring as the coefficient of  $e(\langle Z, v \rangle)$  in the Fourier expansion (3)<sub>A</sub>. For  $\mu, v \neq 0$ , the integral in  $G(r, v)$  may be seen to be  $O((r|v|)^{-N})$  for any  $N \in \mathbb{N}$  with the  $O$ -constant independent of  $\gamma$  while, for  $\mu = 0$ , the integral can be evaluated ([5], (9.10) et seq.) and is well-behaved at infinity as a function of  $r$ . Therefore

the series  $\sum_v G(r, v)e(\langle Z, v \rangle)$  is absolutely convergent for  $\operatorname{Re}(s) > k+1/2$  since, from

[5], we know that the Linnik-Selberg series  $Z(\mu, v, s) := \sum_v \frac{S(\mu, v; \gamma)}{(\gamma \bar{\gamma})^s}$  converges

for  $\text{Re}(s) > k + 1/2$ . Since, for any  $\mu$  in  $\Lambda^\#$ , the growth of  $U_\mu(\cdot, s)$  as  $r$  tends to infinity, is governed by that of  $(\max(r^{\text{Re}(s)}, r^{\text{Re}((k+1/2)-s)})) \times \exp(-2\pi|\mu|r)$ , the inner product  $\langle U_\mu(\cdot, s_1), U_\nu(\cdot, \bar{s}_2) \rangle$  is well-defined for  $\text{Re } s_1, \text{Re } s_2 > k + 1/2$  whenever at least one of  $\mu, \nu$  is not zero. We now proceed to compute this inner product first for  $\mu \neq 0, \nu = 0$ . In fact unfolding the same yields

$$\begin{aligned} \langle U_\mu(\cdot, s_1), E(\cdot, \bar{s}_2) \rangle &= \frac{2\pi^{s_2} |\mu|^{s_2 - (k+1)/2}}{\Gamma(s_2)} \sum_{\gamma \bar{\gamma} \neq 0} \frac{\overline{S(0, \mu; \gamma)}}{(\gamma \bar{\gamma})^{s_2}} \times \\ &\times \int_0^\infty r^{s_1 - (k+3)/2} \exp(-2\pi|\mu|r) K_{s_2 - (k+1)/2}(2\pi|\mu|r) dr \end{aligned}$$

If we now set (formally)  $s_2 = (k+1)/2 - i\tau$  (with  $\tau \in \mathbf{R}$ ) and  $s_1 = k+1 + it$  (with real  $t$ ), then we obtain

$$\begin{aligned} \langle U_\mu(\cdot, k+1+it), E(\cdot, (k+1)/2 + i\tau) \rangle &= \frac{2\pi^{(k+1)/2 - i\tau} |\mu|^{-i\tau}}{\Gamma((k+1)/2 - i\tau)} \times \\ &\times \sum_{\gamma \bar{\gamma} \neq 0} \frac{\overline{S(0, \mu; \gamma)}}{(\gamma \bar{\gamma})^{(k+1)/2 + i\tau}} \int_0^\infty r^{(k-1)/2 + it} \exp(-2\pi|\mu|r) K_{i\tau}(2\pi|\mu|r) dr \\ &= \frac{2\pi^{(k+1)/2 - i\tau} |\mu|^{-i\tau}}{\Gamma\left(\frac{k+1}{2} - i\tau\right)} \times \\ &\times \sum_{\gamma \bar{\gamma} \neq 0} \frac{\overline{S(0, \mu; \gamma)} \sqrt{\pi} \Gamma((k+1)/2 + i(t+\tau)) \Gamma((k+1)/2 + i(t-\tau))}{(\gamma \bar{\gamma})^{(k+1)/2 + i\tau} (4\pi|\mu|)^{(k+1)/2 + i\tau} \Gamma((k+2)/2 + it)}. \end{aligned}$$

since

$$\int_0^\infty \exp(-\alpha x) K_\nu(\alpha x) x^{s-1} dx = \frac{\sqrt{\pi} \Gamma(s+\nu) \Gamma(s-\nu)}{(2\alpha)^s \Gamma(s+1/2)} (\text{Re}(\alpha) > 0, \text{Re}(s) > |\text{Re}(\nu)|).$$

Thus

$$\begin{aligned} \langle U_\mu(\cdot, k+1+it), E(\cdot, (k+1)/2 + i\tau) \rangle &= \\ &= \frac{2\pi^{(k+2)/2 - i\tau} |\mu|^{-i\tau}}{(4\pi|\mu|)^{(k+1)/2 + i\tau}} \frac{\Gamma\left(\frac{k+1}{2} + i(t+\tau)\right) \Gamma\left(\frac{k+1}{2} + i(t-\tau)\right)}{\Gamma\left(\frac{k+1}{2} - i\tau\right) \Gamma\left(\frac{k+2}{2} + it\right)} \\ &\quad \sum \frac{\overline{S(0, \mu; \gamma)}}{(\gamma \bar{\gamma})^{(k+1)/2 - i\tau}} \\ &= \frac{2\pi^{(k+2)/2 - i\tau} |\mu|^{-i\tau}}{(4\pi|\mu|)^{(k+1)/2 + i\tau}} \frac{\Gamma\left(\frac{k+1}{2} + i(t+\tau)\right) \Gamma\left(\frac{k+1}{2} + i(t-\tau)\right)}{\Gamma\left(\frac{k+1}{2} - i\tau\right) \Gamma\left(\frac{k+2}{2} + it\right)} \\ &\quad \bar{\mathbf{Z}}\left(0, \mu, \frac{k+1}{2} + i\tau\right). \end{aligned}$$

We recall that for  $\mu \neq 0$  in  $\Lambda^\#$ ,  $U_\mu(., s)$  has a meromorphic continuation for  $\text{Re}(s) > k$  and is actually holomorphic for  $\text{Re}(s) > k + 1/2$ ; its finitely many poles may come from  $s_l$  in  $\mathbf{R}$  such that  $\mu_l := s_l(k + 1 - s_l)$  is an "exceptional" eigenvalue for  $-\Delta$  i.e.  $0 < s_l \leq k + 1/2$  (cf. [5]). In addition to the eigenfunctions  $v_l$  corresponding to these "exceptional" eigenvalues, the discrete spectrum for  $-\Delta$  in  $L^2(\Gamma \backslash \mathbf{H}^{k+2})$  consisting of  $\{0\} \cup \{\lambda_j | j = 1, 2, \dots\}$  gives rise to corresponding eigenfunctions  $u_0 = \text{constant}$  and  $\{u_1, u_2, \dots\}$  respectively. If we define

$$\tilde{\chi}_l := i\sqrt{(k+1)^2/4 - \mu_l}, \quad \chi_j := \sqrt{\lambda_j - (k+1)^2/4}$$

corresponding respectively to eigenvalues  $\mu_l$  that are "exceptional" and to "non-exceptional" eigenvalues  $\lambda_j$ , then any eigenfunction  $u_\rho$  for  $\rho = i\tilde{\chi}_l$  or  $i\chi_j$  corresponding to  $\mu_l$  or  $\lambda_j$  has the Fourier expansion

$$u_\rho(w) = b_\rho(0)r^{(k+1)/2-s} + \sum_{0 \neq \mu \in \Lambda^\#} a_\rho(\mu)r^{(k+1)/2} K_\rho(2\pi|\mu|r)e(\langle \mu, Z \rangle)$$

with a constant  $b_\rho(0)$  possibly non-zero for  $\rho = i\tilde{\chi}_l$ .

For any eigenfunction  $u_\rho$ , we see that

$$\begin{aligned} \langle u_\rho, U_\mu(., k+1+it) \rangle &= 2\sqrt{\pi}(4\pi|\mu|)^{(k+1/2)-(k+1-it)} v(\mathcal{F}_\Lambda) a_\rho(\mu) \times \\ &\times \frac{\Gamma(k+1-it-(k+1)/2-\rho)\Gamma(k+1-it-((k+1)/2-\rho))}{\Gamma(k+1-it-k/2)}. \end{aligned}$$

By analytic continuation, the Parseval relation now gives us

$$\begin{aligned} \langle U_\mu(., k+1+it), U_\mu(., k+1+it) \rangle &= \\ &= \frac{v^2(\mathcal{F}_\Lambda)}{(4\pi)^k |\mu|^{k+1} |\Gamma(k/2+1+it)|^2} \sum_\rho |a_\rho(\mu)|^2 \left| \Gamma\left(\frac{k+1}{2} - it + \rho\right) \right|^2 \times \\ &\times \left| \Gamma\left(\frac{k+1}{2} + it - \rho\right) \right|^2 + \frac{4\pi^{k+2}}{(4\pi|\mu|)^{k+1}} \times \\ &\times \int_{-\infty}^{\infty} \frac{\left| \Gamma\left(\frac{k+1}{2} + i(t+\tau)\right) \right|^2 \left| \Gamma\left(\frac{k+1}{2} + i(t-\tau)\right) \right|^2}{\left| \Gamma\left(\frac{k+1}{2} - i\tau\right) \right|^2 \left| \Gamma\left(\frac{k+2}{2} + it\right) \right|^2} \\ &\times \left| Z\left(0, \mu, \frac{k+1}{2} + it\right) \right|^2 \frac{d\tau}{4\pi} \end{aligned} \quad (4)_A$$

where  $\rho$  is summed over the non-zero part of the discrete spectrum of  $-\Delta$ . This inner product, on using the Fourier expansion of  $U_\mu(., k+1+it)$  (and unfolding) becomes

$$\int_{\Gamma_\infty \backslash \mathbf{H}^{k+2}} U_\mu(w, k+1+it) r^{k+1-it} \exp(-2\pi|\mu|r - 2\pi i \langle Z, \mu \rangle) \frac{dZ dr}{r^{k+2}} \quad (5)_A$$

$$\begin{aligned}
&= v(\mathcal{F} \wedge) C_{\mu, \mu} \int_0^\infty \exp(-4\pi|\mu|r) r^k dr + \int_0^\infty \frac{r^{k+1-2it}}{r^{k+2}} \exp(-2\pi|\mu|r) dr \times \\
&\times \sum_{\gamma \bar{\gamma} \neq 0} \frac{S(\mu, \mu, \gamma)}{(\gamma \bar{\gamma})^{k+1+it}} \int_{V_{k+1}} \frac{\exp\left(-\frac{2\pi|\mu|}{r|\gamma|^2(1+|Z|^2)}\right)}{(1+|Z|^2)^{k+1+it}} \\
&\exp\left(-\left\langle \frac{(\gamma^*)^{-1} \bar{Z} \gamma^{-1}}{r(1+|Z|^2)}, \mu \right\rangle - r \langle Z, \mu \rangle dZ.
\end{aligned}$$

As in Kuznetsov [3] we now define

$$H_k(\tau, t) := \frac{\left| \Gamma\left(\frac{k+1}{2} + i(t+\tau)\right) \right|^2 \left| \Gamma\left(\frac{k+1}{2} + i(t-\tau)\right) \right|^2}{\pi \left| \Gamma\left(\frac{k+1}{2} - i\tau\right) \right|^2}$$

and for  $\tau \in \mathbb{C}$  with  $|Im(\tau)| < A$  (a constant) and given parameter  $Y > 0$ , we set

$$h_Y(\tau) := \int_0^Y \sin(\pi(\kappa_k + it)) \exp(\pi i(\kappa_k - 1/2)) H_k(\tau, t) dt \quad (6)$$

( $\kappa_k$  being 0 or 1/2 according as  $k$  is even or odd). For real  $\tau$ , we note that  $H_k(\tau, t) = \varphi_k(\tau, t) H_{2\kappa_k}(\tau, t)$  where  $H_0(\tau, t)$  is precisely the kernel  $H(\tau, t)$  in [3],

$$\begin{aligned}
\cosh \pi t H_1(\tau, t) &= \frac{t^2 - \tau^2}{2\tau} \left( \frac{1}{\sinh(\pi(t-\tau))} - \frac{1}{\sinh(\pi(t+\tau))} \right) \text{ and} \\
\varphi_k(\tau, t) &= \begin{cases} \prod_{j=1}^{k/2} \frac{((j+1/2)^2 + (t+\tau)^2)((j+1/2)^2 + (t-\tau)^2)}{((j+1/2)^2 + \tau^2)} & (k \text{ even}) \\ \prod_{j=1}^{(k-1)/2} \frac{(j^2 + (t+\tau)^2)(j^2 + (t-\tau)^2)}{(j^2 + \tau^2)} & (k \text{ odd}) \end{cases}
\end{aligned}$$

Further, for  $\tau$  real and for  $t \geq 0$ , we readily see that

$$\varphi_k(\tau, t) \geq c_k := \begin{cases} \prod_{j=1}^{k/2} (j+1/2)^2 & (k \text{ even}) \\ \prod_{j=1}^{k/2} j^2 & (k \text{ odd}) \end{cases} \quad (7)_A$$

Multiplying both sides of the equation  $(4)_A = (5)_A$  by

$$\frac{(4\pi|\mu|)^{k+1}}{\pi(v(\mathcal{F}))^2} \exp(\pi i(\chi_k - 1/2)) \sin(\pi(\chi_k + it)) \left| \Gamma\left(\frac{k+2}{2} + it\right) \right|^2$$

and then integrating both sides with respect to  $t$  over  $(0, Y)$ , we get

$$\begin{aligned}
& 4 \sum_{\rho} \frac{|a_{\rho}(\mu)|^2}{\left|1/\Gamma\left(\frac{k+1}{2} - \rho\right)\right|^2} h_Y(-i\rho) + \frac{4\pi^{k+1}}{v^2(\mathcal{F}_{\Lambda})} \int_0^Y \exp(\pi i(\chi_k - \frac{1}{2})) \sin(\pi(\chi_k + it)) \times \\
& \times \int_{-\infty}^{\infty} H_k(\tau, t) \left| Z\left(0, \mu, \frac{k+1}{2} + i\tau\right) \right|^2 \frac{d\tau}{4\pi} dt = \frac{1}{\pi} \frac{\Gamma(k+1)}{v(\mathcal{F}_{\Lambda})} C_{\mu, \mu} \int_0^Y P_k(t) dt \\
& + \frac{(4\pi|\mu|)^{k+1}}{\pi v^2(\mathcal{F}_{\Lambda})} \int_0^Y \sum_{\gamma \bar{\gamma} \neq 0} \frac{S(\mu, \mu; \gamma)}{(\gamma \bar{\gamma})^{k+1+i}} P_k(t) dt \int_0^{\infty} \int_{v_{k+1}} r^{k+1-2it} \\
& \times \exp\left(-2\pi|\mu| \left(r + \frac{1}{r|\gamma|^2(1+|Z|^2)}\right)\right) \times e\left(-\left\langle \frac{(\gamma^*)^{-1} \bar{Z} \gamma^{-1}}{r(1+|Z|^2)}, \mu \right\rangle\right) \\
& \times e(-r\langle Z, \mu \rangle) \frac{dr}{r^{k+2}} dZ \tag{8}
\end{aligned}$$

where  $P_k(t) = \frac{1}{\pi} \exp(\pi i) \left(\chi_k - \frac{1}{2}\right) \sin(\pi(\chi_k + it)) \left| \Gamma\left(\frac{k+2}{2} + it\right) \right|^2$  is a monic polynomial in  $t$  of degree  $k+1$ .

We are now led to the estimation of

$$\begin{aligned}
I(\gamma) &:= \int_0^Y \frac{P_k(t) dt}{(\gamma \bar{\gamma})^{it}} \int_0^{\infty} \int_{v_{k+1}} \exp\left[-2\pi \left| \mu \left(r + \frac{1}{r\delta_1}\right) - \right. \right. \\
& \quad \left. \left. - 2\pi i \left\langle \frac{\gamma^{*-1} \bar{Z} \gamma^{-1}}{r(1+|Z|^2)}, \mu \right\rangle + r\langle Z, \mu \rangle \right) \right] \\
& \quad \frac{dr}{r^{1+2it}} \frac{dZ}{(1+|Z|^2)^{k+1+it}} \tag{9}
\end{aligned}$$

with  $\delta_1 := |\gamma|^2(1+|Z|^2)$  and then of the series

$$\sum_{\gamma \bar{\gamma} \neq 0} \frac{S(\mu, \mu; \gamma)}{(\gamma \bar{\gamma})^{k+1}} I(\gamma).$$

We have, for  $k > 0$ ,

$$I(\gamma) \ll Y^{k+3/2} (|\mu|/|\gamma|)^{-1+\beta} \quad \forall \beta \in (0, 1). \tag{10}$$

We now proceed to estimate

$$h_Y(\tau) := \begin{cases} \int_0^Y \sinh \pi t H_k(\tau, t) dt & \text{for } k \text{ even} \\ \int_0^Y \cosh \pi t H_k(\tau, t) dt & \text{for } k \text{ odd} \end{cases} \tag{6'}$$

first for  $\tau$  satisfying  $1 \leq \tau \leq Y - \log Y$  with  $Y$  large. In fact,

$$h_Y(\tau) = h_{\infty}(\tau) - \int_Y^{\infty} \frac{\sinh}{\cosh} \pi t H_k(\tau, t) dt.$$

For odd  $k \geq 1$ ,

$$\int_Y^{\infty} \cosh \pi t H_k(\tau, t) dt = \int_Y^{\infty} \frac{t^2 - \tau^2}{2\tau} \left( \frac{1}{\sinh(\pi(t-\tau))} - \frac{1}{\sinh(\pi(t+\tau))} \right) \varphi_k(\tau, t) dt$$



$$\ll \frac{1}{\tau \prod_{j=1}^{k-1} (j^2 + \tau^2)} \left\{ \int_{Y-\tau}^{\infty} \frac{u(u+2\tau)}{\sinh \pi u} \sum_{l,m=0}^{k-1} u^l (u+2\tau)^m du + \right. \\ \left. + \exp(-\pi\tau) \int_Y^{\infty} \exp(-\pi t) t^{2k-2} dt \right\} \\ \ll \frac{1}{\tau^k} \int_{Y-\tau}^{\infty} u^{2k} \tau^k \exp(-\pi u) + \exp(-\pi\tau) \ll Y^{-(\pi-\varepsilon)} + \exp(-\pi\tau) \\ \text{for every } \varepsilon > 0.$$

The same estimate

$$h_{\infty}(\tau) - h_Y(\tau) \ll Y^{(\pi-\varepsilon)} + \exp(-\pi\tau) \text{ for } 1 \leq \tau \leq Y - \log Y \quad (11)$$

holds also for even  $k \geq 0$ .

On the other hand, for  $\tau \geq Y + \log Y$ , we claim that

$$h_Y(\tau) \ll \exp(-(\pi - \varepsilon)(\tau - Y)) \text{ for every } \varepsilon > 0. \quad (12)$$

It is easy to get lower bounds for  $h_{\infty}(\tau)$  for all  $\tau \geq 1$ . Actually, from (7)<sub>A</sub>,

$$h_{\infty}(\tau) \geq c_k \begin{cases} \int_0^{\infty} \sinh(\pi t) H_0(\tau, t) dt & \text{for } k \text{ even} \\ \int_0^{\infty} \cosh(\pi t) H_1(\tau, t) dt & \text{for } k \text{ odd} \end{cases} \quad (13)_A$$

The integral in (13)<sub>A</sub> for even  $k$  is

$$\geq \frac{1}{2} \int_0^{\infty} \sinh(\pi t) \exp(\pi t) \exp(\pi(t + \tau) - \pi|\tau - t|) dt \geq (1 - e^{-\pi})/2\pi$$

while, for odd  $k$ , the integral equals  $2 \int_0^{\infty} u / \sinh(\pi u) du$  which is clearly independent of  $\tau$ . We thus obtain

$$h_{\infty}(\tau) \geq c'_1 = \min \left( \frac{1 - \exp(-\pi)}{2\pi}, 2 \int_0^{\infty} \frac{u}{\sinh \pi u} du \right) \quad (13)'$$

Now, as before, let us define

$$A(X; \mu) = A(X) := \sum_{|\rho| \leq X} \frac{|a_{\rho}(\mu)|^2}{1 / \left| \Gamma \left( \frac{k+1}{2} - \rho \right) \right|^2}$$

We need only an analogue of Weil's classical estimate for the generalized Kloosterman sums  $S(\mu, \mu; \gamma)$  as well as an appropriate estimate for  $|Z(0, \mu, k + 1/2 + i\tau)|^2$  (due to Elstrodt-Grunewald-Mennicke and unpublished as yet) in order to prove the following formula for  $A(X)$ , carrying out the remaining steps exactly as in the case  $k = 1$ :

For  $k \geq 2$  and any  $\varepsilon > 0$ , we have

$$A(X; \mu) = A(X) = B_k X^{k+2} + O(X^{k+3/2} |\mu|^{k+\varepsilon})$$

as  $X$  tends to infinity, with an explicit constant  $B_k$  depending on  $k$  and the  $O$ -constant depending at most on  $\varepsilon$  and  $k$ .

## References

- [1] Elstrodt J, Grunewald F and Mennicke J, Eisenstein series on three-dimensional hyperbolic space and imaginary quadratic number fields, *J. Reine Angew. Math.* **360** (1985) 160–213
- [2] Gundlach K B, Über die Darstellung der ganzen Spitzenformen zu den Idealstufen der Hilbertschen Modulgruppe und die Abschätzung ihrer Fourierkoeffizienten, *Acta Math.* **92** (1954) 309–345
- [3] Kuznetsov N V, Petersson's conjecture for cusp forms of weight zero and Linnik's conjecture. Sums of Kloosterman sums, *Math. USSR. S.* **39** (1981) 299–342
- [4] Sarnak P, The arithmetic and geometry of some hyperbolic manifolds, *Acta Math.* **151** (1983) 253–295
- [5] Elstrodt J *et al*, Kloosterman sums for Clifford algebras and a lower bound for the positive eigenvalues of the Laplacian for congruence subgroups acting on hyperbolic spaces, *Inv. Math.* **101** (1990) 641–685

## On Zagier's cusp form and the Ramanujan $\tau$ function

ASHWAQ HASHIM and M RAM MURTY

Department of Mathematics, McGill University, Montreal H3A 2K6, Canada

Dedicated to the memory of Professor K G Ramanathan

**Abstract.** Zagier constructed a cusp form for each weight  $k$  of the full modular group. We use this construction to estimate Fourier coefficients of cusp forms. In particular, we get a non-trivial estimate, by elementary methods and indicate a relationship with the Lindelöf hypothesis for classical Dirichlet L-functions.

**Keywords.** Ramanujan tau function; cusp forms.

In [Z], Zagier constructs a non-zero cusp form of weight  $k$  for the full modular group for every even integer  $k \geq 12$ . More precisely, define for real numbers  $\Delta$  and  $r$  satisfying  $\Delta < r^2$  and  $s \in \mathbb{C}$  with  $1/2 < \operatorname{Re}(s) < k$ ,

$$I_k(\Delta, r; s) = \int_0^\infty \int_{-\infty}^\infty \frac{y^{k+s-2} dx dy}{(x^2 + y^2 + iry - \Delta/4)^k}.$$

It is easily seen that if  $\Delta \neq 0$ , we can simplify this to

$$I_k(\Delta, r; s) = \frac{\Gamma(k-1/2)\Gamma(1/2)}{\Gamma(k)} \int_0^\infty \frac{y^{k+s-2} dy}{(y^2 + iry - \Delta/4)^{k-1/2}}.$$

The advantage of the last expression is that the integral is absolutely convergent for  $1-k < \operatorname{Re}(s) < k$ . Moreover, it can be expressed in terms of Legendre functions.

Now let  $\Delta$  be any discriminant. That is  $\Delta \in \mathbb{Z}$  and  $\Delta \equiv 0$  or  $1 \pmod{4}$ . Consider the binary quadratic forms

$$\phi(u, v) = au^2 + buv + cv^2 \quad (a, b, c \in \mathbb{Z})$$

with discriminant  $|\phi| = b^2 - 4ac = \Delta$ . The group  $\Gamma = SL_2(\mathbb{Z})$  operates on the set of such forms by

$$(\gamma\phi)(u, v) = \phi(au + cv, bu + dv)$$

when

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma.$$

$$\zeta(s, \Delta) = \sum_{\substack{\phi \bmod \Gamma \\ |\phi| = \Delta}} \sum_{\substack{(m,n) \in \mathbb{Z}^2 / \Gamma_\phi \\ \phi(m,n) > 0}} \phi(m,n)^{-s},$$

where the first sum is over all  $\Gamma$ -equivalence classes of forms  $\phi$  of discriminant  $\Delta$  and the second is over inequivalent pairs of integers with respect to the stabilizer  $\Gamma_\phi$  of  $\phi$ . If  $\Delta$  is the discriminant of a real or imaginary quadratic field  $K$ , then  $\zeta(s, \Delta) = \zeta(s) L(s, \Delta)$  where  $\zeta(s)$  is the Riemann zeta function and  $L(s, \Delta)$  is the classical Dirichlet series attached to the quadratic character  $(\Delta/\cdot)$ . If  $\Delta = Df^2$  where  $D$  equals 1 or the discriminant of a quadratic field and  $f$  is a natural number, then  $\zeta(s, \Delta)$  differs from  $\zeta(s, D)$  only by a finite Dirichlet series. Therefore, in all cases, we can write

$$\zeta(s, \Delta) = \zeta(s) L(s, \Delta)$$

and this defines  $L(s, \Delta)$ .

Define, for each natural number  $m$ , the function  $F(m, s)$  as equal to zero if  $m$  is not a perfect square and if  $m = u^2, u > 0$ , then

$$F(m, s) = (-1)^{k/2} \frac{\Gamma(k+s-1) \zeta(2s)}{2^{2s+k-3} \pi^s \Gamma(k)} u^{k+s-1}.$$

Then Zagier [Z, p. 110] proved the following:

#### PROPOSITION

For  $m = 1, 2, \dots$  and  $s \in \mathbb{C}$ , set

$$c_m(s) = F(m, s) + m^{k-1} \sum_{r=-\infty}^{\infty} [I_k(r^2 - 4m, r; s) + I_k(r^2 - 4m, -r; s)] L(s, r^2 - 4m).$$

Then,

- (a) the series defining  $c_m(s)$  converges absolutely and uniformly for  $2 - k < \operatorname{Re}(s) < k - 1$ ;
- (b) the function

$$\Phi_s(z) = \sum_{m=1}^{\infty} c_m(s) e^{2\pi i m z}$$

for  $\operatorname{Im}(z) > 0$  and  $2 - k < \operatorname{Re}(s) < k - 1$  is a cusp form of weight  $k$  for the full modular group;

- (c) if  $f$  is a normalized, cuspidal, Hecke eigenform, then

$$(\Phi_s, f) = \frac{(-1)^{k/2} \pi \Gamma(s+k-1)}{2^{k-3} (k-1) (4\pi)^{s+k-1}} D_f(s+k-1)$$

where  $(\cdot, \cdot)$  denotes the Petersson inner product and  $D_f(s)$  is defined as follows. Let

$$f(z) = \sum_{n=1}^{\infty} a(n) e^{2\pi i n z}$$

be the Fourier expansion about the cusp  $i\infty$ . Then,

$$D_f(s) = \frac{\zeta(2s - 2k + 2)}{\zeta(s - k + 1)} \sum_{n=1}^{\infty} \frac{a(n)^2}{n^s}.$$

Thus, Zagier's cusp form has many virtues. We see from (c) that  $\Phi_s$  is not identically zero whenever the space of cusp forms is non-zero and so, the formula for  $c_m(s)$  gives an explicit formula for the Fourier coefficient of a cusp form of weight  $k$ . (In particular, for  $k = 12$ , we get an explicit formula for  $\tau(m)$ .) Indeed, specializing at  $s = 1$  gives the classical Eichler-Selberg trace formula. Moreover, (c) gives the analytic continuation of the "symmetric square  $L$ -function" attached to  $f$ . That is, if  $\pi_f$  is the associated automorphic representation of  $GL_2(\mathbb{A}_{\mathbb{Q}})$ , then  $D_f(s)$  is essentially the Langlands  $L$ -function  $L(s, \text{Sym}^2(\pi_f))$ , which we know is an  $L$ -function attached to a cuspidal automorphic representation of  $GL_3(\mathbb{A}_{\mathbb{Q}})$ , by the work of Gelbart-Jacquet [GJ].

Our purpose here is to indicate that Zagier's cusp form can be used to derive estimates for the Ramanujan  $\tau$ -function (and other coefficients of cusp forms of small weight). Of course, Deligne [D] proved that

$$\tau(n) = O(n^{11/2+\epsilon}).$$

Moreover, we know from [Mu] that this result is best possible. However, the work of Deligne [D] uses deep methods of algebraic geometry. In [L], Langlands suggests an analytic approach which requires analytic continuation of all the symmetric power  $L$ -functions,  $L(s, \text{Sym}^r(\pi_f))$  for a fixed half-plane. At present, such an analytic continuation exists only for  $r \leq 5$  by the work of Hecke [H], Shimura [S] and Shahidi [Sh]. Shahidi's theorems lead to

$$\tau(p) = O(p^{11/2+1/5})$$

for every prime number  $p$ . (It should be stressed that the method of Langlands addresses both holomorphic and non-holomorphic modular forms and not just the holomorphic ones to which the Deligne estimate applies). However, even these results require a great amount of representation theoretic background. Perhaps, the "simplest" method is that of Selberg [Se] using the method of Poincaré series for the explicit construction of cusp forms. His method leads to

$$\tau(n) = O(n^{11/2+1/4+\epsilon}).$$

We believe that Zagier's cusp form has important uses in number theory (and the optimal one not being the use we are going to make below). We will use it to derive a quick estimate for  $\tau(p)$ . We will prove

**Theorem.**  $\tau(p) = O(p^{11/2+7/16+\epsilon}).$

*Remark.* The same estimate is valid for  $\tau(n)$  as well. In fact, the method below is valid *mutatis mutandis* for  $n$  not a perfect square. When  $n$  is a perfect square, the  $F(n, s)$  term is  $O(n^{11/2+1/4+\epsilon})$  for  $s = 1/2 + \delta + it$  and  $\delta > 0$  and arbitrarily small. The bound  $\tau(p) = O(p^6)$  is straightforward and derived in any book on modular forms. We also remark that it should be possible to obtain the Selberg bound by the methods of this paper. Indeed, if we assume the Lindelöf hypothesis for  $L(s, \Delta)$ , it is easily seen

that the Selberg estimate follows. Since this hypothesis is being applied only "on average" rather than to a single  $L$ -function, such a result should follow with a little more technical refinement.

We will need various estimates for the Dirichlet  $L$ -function in the critical strip.

*Lemma 1.*  $L(1/2 + it, \Delta) \ll (|\Delta|(|t| + 2))^{3/16 + \varepsilon}$

*Proof.* For fundamental discriminants and  $t = 0$  this is a result of Burgess [B] and the hybrid estimate is due to Heath-Brown [HB]. So we need only prove this for non-fundamental discriminants. Writing  $\Delta = Df^2$ , where  $D$  is a fundamental discriminant, we see that ([Z, p. 130])

$$L(s, \Delta) = L(s, D) \sum_{d|f} \mu(d) \left(\frac{D}{d}\right) d^{-s} \sigma_{1-2s}(f/d)$$

where

$$\sigma_w(n) = \sum_{d|n} d^w$$

When  $s = 1/2 + it$ , we find

$$L(1/2 + it, \Delta) \ll (|\Delta|(|t| + 2))^{3/16 + \varepsilon}$$

which gives the desired result.

By the Phragmén-Lindelöf principle (see [Ra]) we deduce that:

*Lemma 2.* For  $0 \leq \delta < 1/2$ ,

$$L(1/2 + \delta + it, \Delta) \ll (|\Delta|(|t| + 2))^{(3/16 + \varepsilon)(1 - 2\delta)}.$$

We can now prove the theorem. Choose  $s = 1/2 + \delta$  with  $\delta > 0$  and sufficiently small. Then

$$|c_p(1/2 + \delta)| \leq p^{k-1} \sum_{r=-\infty}^{\infty} |I_k(\Delta, r, 1/2 + \delta) + I_k(\Delta, -r, 1/2 + \delta)| \\ |L(1/2 + \delta, \Delta)|.$$

The integrals can be estimated without difficulty. Indeed,

$$|I_k(\Delta, r, 1/2 + \delta)| \leq \int_0^{\infty} \frac{y^{k-3/2+\delta} dy}{(y^4 + (r^2/2 + 2p)y^2 + \Delta^2/16)^{k/2-1/4}}.$$

This last integral is bounded by

$$\int_0^{\infty} \frac{y^{k-3/2+\delta} dy}{(y^4 + (r^2/2 + 2p)y^2)^{k/2-1/4}},$$

which is readily evaluated upon substituting  $y = (r^2/2 + 2p)^{1/2}u$ , as

$$= (r^2/2 + 2p)^{-k/2+1/4+\delta/2} \int_0^{\infty} u^{\delta-1} (1+u^2)^{-k/2+1/4} du.$$

$$c_p(1/2 + \delta) \ll p^{k-1} \sum_{r=0} (r^2 + 4p)^{-k/2 + 1/4 + \delta/2 + (3/16 + \varepsilon)(1 - 2\delta)}$$

for any  $\delta > 0$ . The series is dominated by the integral

$$\int_0^\infty (x^2 + 4p)^{-v} = \frac{\Gamma(v - 1/2) \Gamma(1/2)}{2\Gamma(v)} (4p)^{1/2 - v},$$

$$v = k/2 - 1/4 - \delta/2 - (3/16 + \varepsilon)(1 - 2\delta)$$

which easily derived by contour integration (or see [GR, p. 295, formula 2]). Setting  $k = 12$ , we deduce

$$\tau(p) = O(p^{6 - 1/16 + \varepsilon}),$$

as desired.

*Remark.* If we use the Lindelöf hypothesis instead of the Burgess-Heath-Brown estimate, we get  $\tau(p) = O(p^{11/2 + 1/4 + \varepsilon})$ . A similar result holds for any cusp form of weight  $k < 24$  or  $k = 26$  for the full modular group because in these cases, the dimension of the space of cusp forms is  $\leq 1$ .

## Acknowledgements

Research work of one of the authors, AH, was partially supported by a scholarship from Bahrain University and the other author, MRM, was partially supported by NSERC and FCAR grants.

## References

- [B] Burgess D, On character sums and  $L$ -series II, *Proc. Lond. Math. Soc.* (3) 13 (1963) 524–536
- [D] Deligne P, La conjecture de Weil I, *Publ. I.H.E.S.* 43 (1974) 273–307
- [GJ] Gelbart S and Jacquet H, A relation between automorphic representations of  $GL(2)$  and  $GL(3)$ , *Ann. Sci. E. Norm. Sup. Ser.* 11 (1978) 471–552
- [H] Hecke E, *Mathematische Werke*, Herausgegeben im Auftrage der Akademie der Wissenschaften zu Göttingen, Vandenhoeck and Ruprecht, Göttingen, (1959) pp. 955
- [L] Langlands R, Problems in the theory of automorphic forms, in *Lectures in Modern Analysis, Springer Lecture Notes* 170 (1970) 18–26
- [GR] Gradshteyn I S and Ryzhik I M, *Tables of integrals, series and products*, 4th edition, Academic Press, (1965)
- [HB] Heath-Brown R, Hybrid bounds for Dirichlet  $L$ -functions II, *Q. J. Math. Oxford* (2) 31 (1980) 157–167
- [Mu] Ram Murty M, On the oscillation of Fourier coefficients of modular forms, *Math. Ann.* 262 (1983) 431–446
- [Ra] Rademacher H, On the Phragmén–Lindelöf theorem and some applications, *Math. Z.* 72 (1959) 192–204
- [Se] Selberg A, On the estimation of Fourier coefficients of modular forms, *Proc. Symp. Pure. Math.* 8 (1965) 1–15

- [S] Shimura G, On the holomorphy of certain Dirichlet series, *Proc. Lond. Math. Soc.*, **31** (1975) 79–98
- [Sh] Shahidi F, On certain  $L$ functions, *Am. J. Math.*, **103** (1981) 297–355
- [Z] Zagier D, Modular forms whose Fourier coefficients involve zeta functions of quadratic fields, in *Modular functions of one variable VI*, (eds. Serre J P and Zagier D B) Lecture Notes in Math., Vol. 627, pp. 105–170, Springer, 1977



# Zeta functions of prehomogeneous vector spaces with coefficients related to periods of automorphic forms

FUMIHIRO SATO

Department of Mathematics, Rikkyo University, Nishi-Ikebukuro, Toshimaku, Tokyo 171, Japan

Dedicated to the memory of Professor K G Ramanathan

**Abstract.** The theory of zeta functions associated with prehomogeneous vector spaces (p.v. for short) provides us a unified approach to functional equations of a large class of zeta functions. However the general theory does not include zeta functions related to automorphic forms such as the Hecke  $L$ -functions and the standard  $L$ -functions of automorphic forms on  $GL(n)$ , even though they can naturally be considered to be associated with p.v.'s. Our aim is to generalize the theory to zeta functions whose coefficients involve periods of automorphic forms, which include the zeta functions mentioned above.

In this paper, we generalize the theory to p.v.'s with symmetric structure of  $K_\pi$ -type and prove the functional equation of zeta functions attached to automorphic forms with generic infinitesimal character. In another paper, we have studied the case where automorphic forms are given by matrix coefficients of irreducible unitary representations of compact groups.

**Keywords.** Zeta function; prehomogeneous vector space; functional equation; automorphic form.

## Introduction

The theory of zeta functions associated with prehomogeneous vector spaces was developed by Sato and Shintani in [S] and [SS] and generalized later in [S1] to multivariable zeta functions. The theory provides us a unified approach to functional equations of a large class of zeta functions. However the general theory does not include zeta functions related to automorphic forms, even though they can naturally be considered to be associated with prehomogeneous vector spaces. Among those are the Hecke  $L$ -functions with Grössencharacters (cf. [Hec1], [Hec2]), the Epstein zeta functions with spherical functions (cf. [E]), the Maass zeta functions attached to quadratic forms (cf. [M1]–[M4]), and the standard  $L$ -functions of automorphic forms on  $GL(n)$  (cf. [GJ]). Our ultimate purpose is to generalize the theory of zeta functions associated with prehomogeneous vector spaces ([SS], [S1]) to zeta functions whose coefficients involve periods of automorphic forms, which include the zeta functions mentioned above.

In [S6], we have studied the case where automorphic forms are given by matrix coefficients of irreducible unitary representations of compact groups. The simplest example in this case is the Epstein zeta functions with spherical functions (cf. § 2.2, [S4]).

In the present paper, we are concerned mainly with prehomogeneous vector spaces with symmetric structure and prove the functional equations and the analytic

infinitesimal character of an automorphic form is generic and the prehomogeneous vector space in question has a symmetric structure of  $K_e$ -type. Our results can be applied, for example, to zeta functions considered in [M3] and [Hej] and some special cases of zeta functions in [M1, 2, 4] as well as several new zeta functions (cf. § 6).

Recall that, in the theory of prehomogeneous vector spaces developed in [SS] and [S1], the proof of the existence of analytic continuations and functional equations of zeta functions is based on the following four properties:

1. Integral representation of the product of zeta functions and local zeta functions at the infinite prime (= the zeta integral);
2. Analytic continuation and the functional equation of the zeta integral;
3. Functional equations satisfied by local zeta functions;
4. The existence of  $b$ -functions (the Bernstein-Sato polynomials), which control the singularities of zeta functions and the gamma-factor of functional equations. Moreover, by using the  $b$ -functions, one can eliminate the troublesome contribution of rational points in the singular set to the zeta integral.

We extend these four properties to zeta functions with automorphic forms.

In § 1, we introduce zeta functions with automorphic forms and give their integral representation (the zeta integral) in a rather general setting. In § 2, as examples of generalized zeta functions defined in § 1, we examine zeta functions with Grössencharacters and the Epstein zeta functions with spherical functions from our point of view. The easiest part of the generalization is the functional equation of the zeta integral, which we describe in § 3. In § 4, we define the notion of symmetric structure of prehomogeneous vector spaces and establish some elementary properties. In § 5, we prove the functional equation of local zeta functions (Theorem 5.3) and that of global zeta functions (Theorem 5.4) under the condition that the infinitesimal character of an automorphic form is generic and its symmetric structure is of  $K_e$ -type. In § 6, we present explicit formulas for the functional equations of zeta functions for some concrete examples of prehomogeneous vector spaces with symmetric structure.

*Notation.* As usual,  $\mathbb{Z}$ ,  $\mathbb{Q}$ ,  $\mathbb{R}$  and  $\mathbb{C}$  stand for the ring of rational integers, the field of rational numbers, the field of real numbers and the field of complex numbers, respectively. For a complex number  $z$ ,  $\operatorname{Re}(z)$  is the real part of  $z$ . Let  $X$  be an affine algebraic variety defined over a field  $K$ . Then the set of  $K$ -rational points of  $X$  is denoted by  $X_K$ . For a linear algebraic group  $G$  defined over a field  $K$ ,

$G^0$  = the identity component of  $G$  with respect to the Zariski topology,

$R_u(G)$  = the unipotent radical of  $G$ ,

$X(G)$  = the group of rational characters of  $G$ ,

$X(G)_K$  = the subgroup of  $X(G)$  of rational characters defined over  $K$ .

For a real Lie group  $G$ ,

$G^0$  = the identity component of  $G$  with respect to the usual Hausdorff topology.

Let  $X$  be a  $C^\infty$ -manifold. Then  $C_0^\infty(X)$  stands for the space of smooth functions on  $X$  with compact support. Let  $V_{\mathbb{R}}$  be a finite-dimensional real vector space. Then  $\mathcal{S}(V_{\mathbb{R}})$  stands for the space of rapidly decreasing functions on  $V_{\mathbb{R}}$ . Let  $V_{\mathbb{Q}}$  be a finite-dimensional  $\mathbb{Q}$ -vector space. A  $\mathbb{C}$ -valued function  $f$  on  $V_{\mathbb{Q}}$  is called a *Schwartz-*

*Bruhat* function, if there exist lattices  $L_1, L_2$  in  $V_{\mathbb{Q}}$  such that the support of  $f$  is contained in  $L_1$  and the value of  $f(x)$  is determined by the residue class  $x \pmod{L_2}$ . We denote by  $\mathcal{S}(V_{\mathbb{Q}})$  the space of Schwartz-Bruhat functions on  $V_{\mathbb{Q}}$ . For  $m, n \geq 1$ , we put

$\mathbf{M}(m, n)$  = the space of  $m$  by  $n$  complex matrices,

$\mathbf{M}(n) = \mathbf{M}(n, n)$  = the ring of complex square matrices of size  $n$ ,

$I_n \in \mathbf{M}(n)$  = the identity matrix of size  $n$ ,

$0^{(m, n)}$  = the zero matrix in  $\mathbf{M}(m, n)$ .

For a subring  $R$  of  $\mathbb{C}$ , denote by  $\mathbf{M}(m, n; R)$  the subset of  $\mathbf{M}(m, n)$  of matrices with entries in  $R$ . For a real Lie algebra  $\mathfrak{g}$ , we denote by  $U(\mathfrak{g})$  the universal enveloping algebra of the complexification  $\mathfrak{g}_{\mathbb{C}} = \mathfrak{g} \otimes_{\mathbb{R}} \mathbb{C}$ .

## 1 Definition of zeta functions and their integral representations

**1.1** Let  $(G, \rho, V)$  be a prehomogeneous vector spaces defined over a field  $K$  of characteristic 0 and denote its singular set by  $S$ . Then, by definition,  $V_{\bar{K}} - S_{\bar{K}}$  is a single  $G_{\bar{K}}$ -orbit, where  $\bar{K}$  is the algebraic closure of  $K$ .

Let  $S_1, \dots, S_n$  be the  $K$ -irreducible hypersurfaces contained in  $S$  and take  $K$ -irreducible polynomials  $P_1, \dots, P_n$  defining  $S_1, \dots, S_n$ , respectively. It is known that the polynomial  $P_i$  is unique up to a non-zero constant factor in  $K$ . For each  $i = 1, \dots, n$ , there exists a  $K$ -rational character  $\chi_i$  satisfying

$$P_i(\rho(g)x) = \chi_i(g)P_i(x) \quad (g \in G, x \in V).$$

We call  $P_1, \dots, P_n$  as the basic relative invariants over  $K$ . Any relative invariant of  $(G, \rho, V)$  with coefficients in  $K$  can be expressed as a product of  $P_1, \dots, P_n$ , negative power being allowed.

Denote by  $X_{\rho}(G)_K$  the subgroup of  $X(G)_K$  generated by  $\chi_1, \dots, \chi_n$ , which is a free abelian group of rank  $n$ . (For basic definitions and properties in the theory of prehomogeneous vector spaces, refer to [S1, § 1] and [SK].)

**1.2** In the following, we consider a prehomogeneous vector space  $(G, \rho, V)$  defined over the rational number field  $\mathbb{Q}$ . For an  $x \in V$ , put

$$G_x = \{g \in G \mid \rho(g)x = x\}.$$

We assume that

- (A-1) for any  $x \in V_{\mathbb{Q}} - S_{\mathbb{Q}}$ , the isotropy subgroup  $G_x$  is reductive and  $X((G_x)^0)_{\mathbb{Q}} = \{1\}$ ;
- (A-2)  $G$  has a semidirect product decomposition  $G = LU$ , where  $L$  is a connected reductive  $\mathbb{Q}$ -subgroup and  $U$  is a connected normal  $\mathbb{Q}$ -subgroup with  $X(U) = \{1\}$ .

The group  $G$  always has a semi-direct product decomposition satisfying (A-2). Namely  $G$  is a semi-direct product of  $U = R_u(G)$ , the unipotent radical, and a Levi subgroup  $L$ . In the following we fix a decomposition  $G = LU$  satisfying (A-2) once for all, which may not be the Levi decomposition (for concrete examples, see § 2, § 4.1 and § 6).

One of the consequences of assumption (A-1) is the following:

*Lemma 1.1* *The singular set  $S$  is a hypersurface.*

*Proof.* Since  $G_x$  is reductive, the open orbit  $V - S$  is an affine variety. This implies that  $S$  is a hypersurface. ■

Let  $P_1, \dots, P_n$  be the basic relative invariants over  $\mathbb{Q}$  and  $\chi_1, \dots, \chi_n$  the rational characters of  $G$  corresponding to  $P_1, \dots, P_n$ , respectively. Let  $G_0$  be the identity component of  $\bigcap_{i=1}^n \ker \chi_i$  with respect to the Zariski topology. Put  $L_0 = L \cap G_0$ . Then  $L_0$  is connected and we have  $G_0 = L_0 U$  (semi-direct product).

*Lemma 1.2* *The groups  $X(L_0)_{\mathbb{Q}}$  and  $X(G_0)_{\mathbb{Q}}$  are trivial.*

*Proof.* By (A-1), (A-2) and [S1, Lemma 4.1], we have

$$\text{rank } X_{\rho}(G)_{\mathbb{Q}} = \text{rank } X(G)_{\mathbb{Q}} = \text{rank } X(L)_{\mathbb{Q}}.$$

This implies that

$$\text{rank } X(G_0)_{\mathbb{Q}} = \text{rank } X(L_0)_{\mathbb{Q}} = 0.$$

Since  $G_0$  and  $L_0$  are connected, the groups  $X(L_0)_{\mathbb{Q}}$  and  $X(G_0)_{\mathbb{Q}}$  are trivial. ■

Let  $T$  be the largest  $\mathbb{Q}$ -split torus of the identity component of the center  $Z(L)$  of  $L$ . Then  $\dim T = \text{rank } X(G)_{\mathbb{Q}} = \text{rank } X_{\rho}(G)_{\mathbb{Q}}$  and  $L$  is an almost direct product of  $T$  and  $L_0$ .

**1.3** Let  $G^+$ ,  $G_0^+$ ,  $T^+$ ,  $L_0^+$  and  $U^+$  be the identity components of the real Lie groups  $G_{\mathbb{R}}$ ,  $G_{0,\mathbb{R}}$ ,  $T_{\mathbb{R}}$ ,  $L_{0,\mathbb{R}}$  and  $U_{\mathbb{R}}$ , respectively. Then we have

$$G^+ = T^+ L_0^+ U^+, \quad G_0^+ = L_0^+ U^+$$

and the decomposition

$$g = thu \quad (g \in G^+, t \in T^+, h \in L_0^+, u \in U^+)$$

is unique. By (A-2) and Lemma 1.2, the groups  $L_0^+$  and  $U^+$  are unimodular.

Let  $dt, dh$  and  $du$  be (bi-invariant) Haar measures on  $T^+$ ,  $L_0^+$  and  $U^+$ , respectively. Let  $d_r g$  be a right invariant measure on  $G^+$  and let  $\Delta: G^+ \rightarrow \mathbb{R}_+^*$  be the module of  $d_r g$ . Then we can normalize these measures so that

$$d_r g = d_r(thu) = \Delta(t) dt dh du.$$

As proved in [S1, §4], the assumption (A-1) ensures the existence of  $\delta = (\delta_1, \dots, \delta_n) \in \mathbb{Q}^n$ , for which

$$\Omega(x) = \prod_{i=1}^n |P_i(x)|^{-\delta_i} dx, \quad dx = \text{the Lebesgue measure on } V_{\mathbb{R}}$$

gives a relatively  $G^+$ -invariant measure on  $V_{\mathbb{R}} - S_{\mathbb{R}}$  with multiplier  $\Delta$ .

$$V_R - S_R = V_1 \cup \dots \cup V_v$$

be the decomposition into connected components. Each connected component  $V_i$  is a single  $G^+$ -orbit. For an  $x \in V_i$ , put  $G_x^+ = G_x \cap G^+$ . By (A-1), the group  $G_x^+$  is a unimodular Lie group. We normalize a (bi-invariant) Haar measure  $d\mu_x$  on  $G_x^+$  such that

$$\int_{G^+} F(g) d_r g = \int_{V_i} \Omega(\rho(\dot{g})x) \int_{G_x^+} F(\dot{g}h) d\mu_x(h) \quad (F \in L^1(G^+, d_r g)). \quad (1.1)$$

**1.4** Let  $\phi: L_0^+ \rightarrow W$  be a function on  $L_0^+$  with values in a finite-dimensional complex vector space  $W$ , which is invariant under the right multiplication of some arithmetic subgroup of  $L_{0,Q} \cap L_0^+$ . Later we shall assume that  $\phi$  is an automorphic form on  $L_0^+$ ; however at the moment we do not assume it.

Now let us associate to  $\phi$  a linear form  $Z_\phi(s)$  on  $\mathcal{S}(V_R) \otimes \mathcal{S}(V_Q)$  with complex parameter  $s$  in  $\mathbb{C}^n$ , which we call the zeta integral attached to  $\phi$  (for the definition of  $\mathcal{S}(V_R)$  and  $\mathcal{S}(V_Q)$ , see Notation. See also [S5, §4]).

Consider the canonical surjection  $p: G_0 \rightarrow L_0 = G_0/U$ . The map  $p$  induces a real analytic mapping

$$p: G_0^+ \rightarrow L_0^+ = G_0^+/U^+.$$

For an arithmetic subgroup  $\Gamma$  of  $G_{0,Q} \cap G_0^+$ , put  $\Gamma_L = p(\Gamma) \subset L_0^+$ . Then  $\Gamma_L$  is an arithmetic subgroup of  $L_{0,Q} \cap L_0^+$  (cf. [Bo, Theorem 6]).

For  $f_\infty \otimes f_0 \in \mathcal{S}(V_R) \otimes \mathcal{S}(V_Q)$ , take an arithmetic subgroup  $\Gamma$  of  $G_{0,Q} \cap G_0^+$  such that  $f_0$  is  $\Gamma$ -invariant and  $\phi$  is  $\Gamma_L$ -invariant. Then we define the zeta integral attached to  $\phi$  by setting

$$\begin{aligned} Z_\phi(s)(f_\infty \otimes f_0) &= Z_\phi(s_1, \dots, s_n)(f_\infty \otimes f_0) \\ &= \frac{1}{v(\Gamma)} \int_{T^+} \prod_{i=1}^n \chi_i(t)^{s_i} \Delta(t) dt \int_{G_0^+/\Gamma} \phi(h) \sum_{x \in V_Q - S_Q} f_0(x) f_\infty(\rho(thu)x) dh du, \end{aligned} \quad (1.2)$$

where  $v(\Gamma) = \int_{G_0^+/\Gamma} dh du$ , which is finite by Lemma 1.2 and [BH, Theorem 9-4]. Note that the integral  $Z_\phi(s)$  is independent of the choice of  $\Gamma$ .

In the following, we assume that

(A-3) for any  $f_\infty \otimes f_0 \in \mathcal{S}(V_R) \otimes \mathcal{S}(V_Q)$ , the integral  $Z_\phi(s)(f_\infty \otimes f_0)$  is absolutely convergent, when  $\text{Re}(s_1), \dots, \text{Re}(s_n)$  are sufficiently large.

In case  $\phi$  is a constant function, the integral  $Z_\phi(s)(f_\infty \otimes f_0)$  gives an integral representation of the usual zeta functions associated with  $(G, \rho, V)$  (see [S1, §4], [S5, §4], [SS, §2]). In this case some sufficient conditions for (A-3) are known by [S2, Theorem 1] and [SS, Lemmas 2.2, 2.5]. For example, we have the following criterion of convergence of  $Z_\phi(s)$ :

### PROPOSITION 1.3

Assume that  $X_\rho(\mathbf{G})_{\mathbb{Q}} = X_\rho(\mathbf{G})_{\mathbb{C}}$ . If  $\mathbf{G}_{0,x} = \mathbf{G}_0 \cap \mathbf{G}_x (x \in \mathbf{V} - \mathbf{S})$  is a connected semisimple algebraic group and  $\phi: L_0^+ \rightarrow W$  is bounded, then  $Z_\phi(f_\infty \otimes f_0)(f_\infty \otimes f_0 \in \mathcal{S}(\mathbf{V}_{\mathbb{R}}) \otimes \mathcal{S}(\mathbf{V}_{\mathbb{Q}}))$  is absolutely convergent for  $\text{Re}(s_1) > \delta_1, \dots, \text{Re}(s_n) > \delta_n$ .

*Proof.* Proposition is an immediate consequence of [S2, Theorem 1] and the recent results of Kottwitz [K] and Chernousov [C]. ■

### COROLLARY 1.4

Assume that  $X_\rho(\mathbf{G})_{\mathbb{Q}} = X_\rho(\mathbf{G})_{\mathbb{C}}$ . If  $\mathbf{G}_{0,x} = \mathbf{G}_0 \cap \mathbf{G}_x (x \in \mathbf{V} - \mathbf{S})$  is a connected semisimple algebraic group and  $\phi$  is a cusp form on  $L_0^+$ , then  $Z_\phi(f_\infty \otimes f_0)(f_\infty \otimes f_0 \in \mathcal{S}(\mathbf{V}_{\mathbb{R}}) \otimes \mathcal{S}(\mathbf{V}_{\mathbb{Q}}))$  is absolutely convergent for  $\text{Re}(s_1) > \delta_1, \dots, \text{Re}(s_n) > \delta_n$ .

**1.5** What we must do first is to find a good condition under which the integral  $Z_\phi(f_\infty \otimes f_0)$  can be decomposed into product of Dirichlet series (related only to  $f_0$ ) and local zeta functions (related only to  $f_\infty$ ), as in the case where  $\phi$  is a constant function (cf. [SS, Lemma 2.5], [S1, Lemma 4.2]).

For an  $x \in \mathbf{V}_{\mathbb{Q}} - \mathbf{S}_{\mathbb{Q}}$ , put  $\Gamma_x = \Gamma \cap G_x^+$ . By (A-1) and [BH, Theorem 9-4], the volume  $\mu(x) = \int_{G_x^+/\Gamma_x} d\mu_x$  is finite (for  $d\mu_x$ , see (1.1)). Also put

$$\begin{aligned} L_{(x)}^+ &= p(G_x^+)(\subset L_0^+), & \Gamma_{(x)} &= p(\Gamma_x)(\subset L_{(x)}^+), \\ U_x^+ &= G_x^+ \cap U^+, & \Gamma_{U,x} &= \Gamma_x \cap U^+. \end{aligned}$$

Here we note that  $G_x^+ \subset G_0^+$ . We normalize Haar measures  $dv_x$  and  $d\tau_x$  on  $L_{(x)}^+$  and  $U_x$ , respectively by

$$\int_{L_{(x)}^+/\Gamma_{(x)}} dv_x = 1 \text{ and } \int_{U_x^+/\Gamma_{U,x}} d\tau_x = \mu(x).$$

Then we have  $d\mu_x = dv_x d\tau_x$  on  $G_x^+$ .

For each connected component  $V_i$  of  $\mathbf{V}_{\mathbb{R}} - \mathbf{S}_{\mathbb{R}}$ , we fix a representative  $x_i$  and put  $X_i = L_0^+/L_{(x_i)}^+$ . For each  $x \in V_i$ , choose  $t_x \in T^+$ ,  $h_x \in L_0^+$  and  $u_x \in U^+$  such that  $x = \rho(t_x h_x u_x) x_i$ . Define a mapping  $\bar{\cdot}: V_i \rightarrow X_i$  by  $x \mapsto \bar{x} = h_x \cdot L_{(x_i)}^+ \in X_i$ . The point  $\bar{x}$  is independent of the choice of  $h_x$  and the mapping  $\bar{\cdot}$  defines a real analytic mapping equivariant under the action of  $L_0^+$ .

For  $x \in \mathbf{V}_{\mathbb{Q}} \cap V_i$  and  $y \in V_i$ , set

$$\mathcal{M}_x^{(i)} \phi(\bar{y}) = \int_{L_{(x)}^+/\Gamma_{(x)}} \phi(h_y h_x^{-1} \eta) dv_x(\eta), \quad (1.3)$$

which we call the mean value of  $\phi$  at  $x$ . We consider  $\mathcal{M}_x^{(i)} \phi$  as a function on  $X_i$ . Now it is easy to check the following lemma:

**Lemma 1.5.** *If  $\text{Re}(s_1), \dots, \text{Re}(s_n)$  are sufficiently large to ensure the absolute convergence*

$$Z_\phi(s)(f_\infty \otimes f_0) =$$

$$\frac{1}{v(\Gamma)} \sum_{i=1}^v \sum_{x \in \Gamma \backslash V_0 \cap V_i} \frac{\mu(x) f_0(x)}{\prod_{j=1}^n |P_j(x)|^{s_j}} \int_{V_i} \prod_{j=1}^n |P_j(y)|^{s_j} f_\infty(y) \mathcal{M}_x^{(i)} \phi(\bar{y}) \Omega(y).$$

**1.6** From now on, we assume that  $\phi$  is an automorphic form on  $L_0^+$  with respect to some arithmetic subgroup. To be precise, let  $K$  be a maximal compact subgroup of  $L_0^+$  and  $\pi$  an irreducible unitary representation of  $K$  on a finite dimensional Hilbert space  $W_\pi$ . Denote by  $\mathcal{Z}(L_0^+)$  the algebra of bi-invariant differential operators on  $L_0^+$ . Let  $\chi: \mathcal{Z}(L_0^+) \rightarrow \mathbb{C}$  be an infinitesimal character. Then we call a function  $\phi: L_0^+ \rightarrow W_\pi$  an automorphic form of type  $(\chi, \pi)$  with respect to  $\Gamma_L$  if it satisfies the conditions

$$D\phi = \chi(D)\phi \quad (D \in \mathcal{Z}(L_0^+)),$$

$$\phi(kh\gamma) = \pi(k)\phi(h) \quad (k \in K, h \in L_0^+, \gamma \in \Gamma_L),$$

$\phi$  is slowly increasing.

We denote by  $\mathcal{A}(L_0^+/\Gamma_L; \chi, \pi)$  the space of automorphic forms of type  $(\chi, \pi)$  with respect to  $\Gamma_L$ . It is known that the dimension of  $\mathcal{A}(L_0^+/\Gamma_L; \chi, \pi)$  is finite ([BJ, Theorem 1.7], [H, Theorem 1]).

Any element  $D \in \mathcal{Z}(L_0^+)$  induces an  $L_0^+$ -invariant differential operator on the homogeneous space  $X_i = L_0^+/L_{(x_i)}^+$ , which we denote by  $\bar{D}$ . We call a function  $\psi: X_i \rightarrow W_\pi$  a spherical function of type  $(\chi, \pi)$  if it satisfies the conditions

$$\bar{D}\psi = \chi(D)\psi \quad (D \in \mathcal{Z}(L_0^+)),$$

$$\psi(k\bar{x}) = \pi(k)\psi(\bar{x}) \quad (k \in K, \bar{x} \in X_i).$$

We denote by  $\mathcal{E}(X_i; \chi, \pi)$  the space of spherical functions of type  $(\chi, \pi)$  on  $X_i$ .

**Lemma 1.6.** *Let  $\phi$  be an automorphic form in  $\mathcal{A}(L_0^+/\Gamma_L; \chi, \pi)$ . If the integral (1.3) converges absolutely, then the mean value  $\mathcal{M}_x^{(i)} \phi$  at  $x$  is in  $\mathcal{E}(X_i; \chi, \pi)$ .*

Our final assumption in this section is the following:

(A-4) the dimension of  $\mathcal{E}(X_i; \chi, \pi)$  ( $1 \leq i \leq v$ ) is finite.

Put  $m_i = \dim \mathcal{E}(X_i; \chi, \pi)$  ( $1 \leq i \leq v$ ) and take a basis  $\{\psi_1^{(i)}, \dots, \psi_{m_i}^{(i)}\}$  of  $\mathcal{E}(X_i; \chi, \pi)$ . By Lemma 1.6, we can express  $\mathcal{M}_x^{(i)} \phi$  as a linear combination of  $\psi_1^{(i)}, \dots, \psi_{m_i}^{(i)}$ :

$$\mathcal{M}_x^{(i)} \phi = \sum_{l=1}^{m_i} c_l^{(i)}(\phi; x) \psi_l^{(i)}. \quad (1.4)$$

The coefficients  $c_l^{(i)}(\phi; x)$  can be viewed as functions of  $x$  on  $\Gamma \backslash V_0 \cap V_i$ .

We define (global) zeta functions  $\zeta_l^{(i)}(\phi, f_0; s)$  and local zeta functions  $\Phi_l^{(i)}(f_\infty; \pi, \chi, s)$

by

$$\zeta_l^{(i)}(\phi, f_0; s) = \frac{1}{v(\Gamma)} \sum_{x \in \Gamma \backslash V_0 \cap V_l} \frac{\mu(x) f_0(x) c_l^{(i)}(\phi; x)}{\prod_{j=1}^n |P_j(x)|^{s_j}},$$

$$\Phi_l^{(i)}(f_\infty; \pi, \chi, s) = \int_{V_l} \prod_{j=1}^n |P_j(y)|^{s_j} \psi_l^{(i)}(\bar{y}) f_\infty(y) \Omega(y) \quad (1 \leq i \leq v, 1 \leq l \leq m_i).$$

The zeta functions  $\zeta_l^{(i)}(\phi, f_0; s)$  are independent of the choice of  $\Gamma$ . By Lemma 1.5 and the identity (1.4), we easily obtain the following:

### PROPOSITION 1.7

Assume that  $(\mathbf{G}, \rho, \mathbf{V})$  satisfies (A-1)–(A-4). Then the following identity holds for sufficiently large  $\text{Re}(s_1), \dots, \text{Re}(s_n)$ :

$$Z_\phi(s)(f_\infty \otimes f_0) = \sum_{i=1}^v \sum_{l=1}^{m_i} \zeta_l^{(i)}(\phi, f_0; s) \Phi_l^{(i)}(f_\infty; \pi, \chi, s).$$

*Remark.* The coefficients  $c_l^{(i)}(\phi; x)$  can be expressed as a linear combination of functions of  $x$  of the form  $(\mathcal{M}_x^{(i)} \phi(\bar{y}_l), e_s)$ , where  $\{\bar{y}_l\}$  are a finite number of points in  $X_l$ ,  $\{e_s\}$  is an orthonormal basis of  $W_\pi$  and  $(\cdot)$  is the inner product on  $W_\pi$ . Thus the coefficients of our zeta functions are, roughly speaking, mean values (or periods) of automorphic forms.

The simplest case where the assumption (A-4) is satisfied is the following:

The case of Grössencharacters —  $\phi$  is a quasi-character of  $L_0^+$ .

It is known that (A-4) holds also in the following two cases:

Compact case —  $L_0^+$  is a compact Lie group (by the theorem of Peter-Weyl);

Symmetric case —  $X_i$  ( $1 \leq i \leq v$ ) are reductive symmetric spaces (by a theorem of van den Ban, see. [B1, Cor. 3.10], [B2, Lemma 2.1]).

## 2. Examples of zeta functions

In this section we give some examples of zeta functions introduced in §1 for the case of Grössencharacters, and for a certain compact case. Examples in symmetric case will be given in §6 after establishing a general theory for symmetric case. We keep the notation in §1.

### 2.1 The case of quasi-characters

Let us consider the case where the automorphic form  $\phi$  is a quasi-character of  $L_0^+$ . Put

$$r_1 = 2 \text{rank } X_\rho(\mathbf{G})_{\mathbb{R}} - \text{rank } X_\rho(\mathbf{G})_{\mathbb{C}} \text{ and } r_2 = \text{rank } X_\rho(\mathbf{G})_{\mathbb{C}} - \text{rank } X_\rho(\mathbf{G})_{\mathbb{R}}.$$

Let  $Q_1, \dots, Q_{r_1+2r_2}$  be the basic relative invariants over  $\mathbb{C}$ . We may assume that

$$Q_1, \dots, Q_{r_1} \in \mathbb{R}[V], \quad \overline{Q_{r_1+i}} = Q_{r_1+r_2+i} \quad (1 \leq i \leq r_2)$$



and the basic relative invariants over  $\mathbb{R}$  are given by

$$Q_1, \dots, Q_{r_1}, Q_{r_1+1} Q_{r_1+r_2+1}, \dots, Q_{r_1+r_2} Q_{r_1+2r_2}.$$

Let  $\xi_1, \dots, \xi_{r_1+2r_2}$  be the rational characters of  $\mathbf{G}$  corresponding to  $Q_1, \dots, Q_{r_1+2r_2}$ , respectively.

**Lemma 2.1.** *If  $\phi$  is a quasi-character of  $L_0^+$  and the zeta integral  $Z_\phi(s)(f_\infty \otimes f_0)$  does not vanish, then  $\phi$  is of the form*

$$\begin{aligned} \phi(h) &= \prod_{i=1}^{r_1+2r_2} \xi_i(h)^{\alpha_i} \\ &= \prod_{i=1}^{r_1} \xi_i(h)^{\alpha_i} \prod_{j=1}^{r_2} |\xi_{r_1+j}(h)|^{\alpha_{r_1+j} + \alpha_{r_1+r_2+j}} \cdot \left( \frac{\xi_{r_1+j}(h)}{|\xi_{r_1+j}(h)|} \right)^{\alpha_{r_1+j} - \alpha_{r_1+r_2+j}} \quad (h \in L_0^+), \end{aligned}$$

where  $\alpha_1, \dots, \alpha_{r_1+2r_2} \in \mathbb{C}$  and  $\alpha_{r_1+j} \equiv \alpha_{r_1+r_2+j} \pmod{\mathbb{Z}}$  ( $1 \leq j \leq r_2$ ).

*Proof.* Since  $\phi$  is a quasi-character, we have

$$\mathcal{M}_x^{(i)} \phi(\bar{y}) = \phi(h_y) \phi(h_x)^{-1} \int_{L_{(x)}^+ / \Gamma_{(x)}} \phi(\eta) d\nu_x(\eta).$$

Therefore, if  $Z_\phi(s)(f_\infty \otimes f_0)$  does not vanish, the quasi-character  $\phi$  is trivial on  $L_{(x)}^+$ . We extend  $\phi$  to a quasi-character of  $G^+$  by  $\phi(t) = \phi(u) = 1$  for  $t \in T^+$  and  $u \in U^+$ . Then  $\ker \phi$  contains  $G_x^+ [L_0^+, L_0^+] U^+$ . Denote by  $\tilde{\mathbf{T}}$  the identity component of the center of  $\mathbf{L}$ . Let  $\tilde{T}^+$  be the identity component of the real Lie group  $\tilde{\mathbf{T}}_{\mathbb{R}}$ . Put

$$\tilde{\mathbf{T}}_{(x)} = \tilde{\mathbf{T}} \cap \mathbf{G}_x[\mathbf{L}, \mathbf{L}] \mathbf{U} \text{ and } \tilde{T}_{(x)}^+ = \tilde{T}^+ \cap \mathbf{G}_x[\mathbf{L}, \mathbf{L}] \mathbf{U}.$$

Note that  $\tilde{T}_{(x)}^+ = \tilde{T}^+ \cap G_x^+ [L_0^+, L_0^+] U^+$ . Hence the quasi-character  $\phi$  induces a quasicharacter of  $\tilde{T}^+ / \tilde{T}_{(x)}^+$ . By [SK, Proposition 19], we have an isomorphism

$$\begin{aligned} \tilde{\mathbf{T}} / \tilde{\mathbf{T}}_{(x)} &\xrightarrow{\cong} \overbrace{\mathbf{GL}(1) \times \dots \times \mathbf{GL}(1)}^{r_1+2r_2} \\ t &\mapsto (\xi_1(t), \dots, \xi_{r_1+2r_2}(t)) \end{aligned}$$

and, for the group of real points, we have

$$\begin{aligned} (\tilde{\mathbf{T}} / \tilde{\mathbf{T}}_{(x)})_{\mathbb{R}} &\xrightarrow{\cong} \overbrace{\mathbb{R}^\times \times \dots \times \mathbb{R}^\times}^{r_1} \times \overbrace{\mathbb{C}^\times \times \dots \times \mathbb{C}^\times}^{r_2} \\ t &\mapsto (\xi_1(t), \dots, \xi_{r_1+r_2}(t)). \end{aligned}$$

Since

$$\tilde{T}^+ / \tilde{T}_{(x)}^+ \xrightarrow{\cong} \overbrace{\mathbb{R}_+^\times \times \dots \times \mathbb{R}_+^\times}^{r_1} \times \overbrace{\mathbb{C}^\times \times \dots \times \mathbb{C}^\times}^{r_2},$$

we see that the quasi-character  $\phi$  is of the form given in Lemma. ■

We decompose the set of the indices  $\{1, 2, \dots, r_1 + 2r_2\}$  as follows:

$$\{1, 2, \dots, r_1 + 2r_2\} = \bigcup_{j=1}^n I_j, \quad \chi_j = \prod_{i \in I_j} \xi_i.$$

Since  $L_0^+$  is contained in the kernel of every character in  $X_\rho(G)_{\mathbb{Q}}$ , we may choose the exponents  $\alpha_1, \dots, \alpha_{r_1+2r_2}$  so that

$$\sum_{i \in I_j} \alpha_i = 0 \quad (1 \leq j \leq n).$$

Put

$$Q^\alpha(x) = \prod_{i=1}^{r_1} |Q_i(x)|^{\alpha_i} \prod_{j=1}^{r_2} |Q_{r_1+j}(x)|^{\alpha_{r_1+j} + \alpha_{r_1+r_2+j}} \cdot \left( \frac{Q_{r_1+j}(x)}{|Q_{r_1+j}(x)|} \right)^{\alpha_{r_1+j} - \alpha_{r_1+r_2+j}} \\ (x \in V_{\mathbb{R}} - S_{\mathbb{R}}).$$

By Lemma 2.1 above and (1.4), it is easy to check the following lemma.

*Lemma 2.2. We have*

$$\mathcal{M}_x^{(i)} \phi(\bar{y}) = Q^\alpha(y)/Q^\alpha(x) \quad \text{for } x \in V_{\mathbb{Q}} \cap V_i \text{ and } y \in V_i.$$

*In particular,  $m_i = 1$  and*

$$c^{(i)}(\phi; x) = Q^\alpha(x)^{-1}, \quad \psi^{(i)}(\bar{y}) = Q^\alpha(y).$$

Put

$$\zeta^{(i)}(\alpha, f_0; s) = \frac{1}{v(\Gamma)} \sum_{x \in \Gamma \backslash V_{\mathbb{Q}} \cap V_i} \frac{\mu(x) f_0(x)}{Q^\alpha(x) \prod_{j=1}^n |P_j(x)|^{s_j}}$$

and

$$\Phi^{(i)}(f_\infty; \alpha, s) = \int_{V_i} \prod_{j=1}^n |P_j(y)|^{s_j} Q^\alpha(y) f_\infty(y) \Omega(y).$$

Then the following proposition is an immediate consequence of Lemma 2.2.

### PROPOSITION 2.3

*Assume that  $(G, \rho, V)$  satisfies (A-1)–(A-3) and  $\phi$  is a quasi-character. Then the following identity holds for sufficiently large  $\operatorname{Re}(s_1), \dots, \operatorname{Re}(s_n)$ :*

$$Z_\phi(s)(f_\infty \otimes f_0) = \sum_{i=1}^v \zeta^{(i)}(\alpha, f_0; s) \Phi^{(i)}(f_\infty; \alpha, s).$$

*Remarks.* (1) Let  $K$  be an algebraic number field and  $V(1)$  the 1-dimensional vector space over  $K$ . Consider the prehomogeneous vector space

$$(G, \rho, V) = R_{K/\mathbb{Q}}(\mathbf{GL}(1), \rho_1, V(1)),$$

where  $R_{K/\mathbb{Q}}$  is Weil's functor of restriction of scalars and  $\rho_1$  is the standard

representation of  $\mathrm{GL}(1)$  on  $V(1)$ . Then we have

$$n = \mathrm{rank} X_\rho(\mathbf{G})_{\mathbb{Q}} = 1, \quad r_1 + 2r_2 = [K:\mathbb{Q}],$$

$r_1$  = the number of real places,  $r_2$  = the number of complex places,

and the zeta functions  $\zeta^{(i)}(\alpha, f_0; s)$  are the Hecke  $L$ -functions with Grössencharacters of  $K$  (cf. [Hec1], [Hec2]).

(2) As we remarked in the introduction, the functional equation of  $\zeta^{(i)}(\alpha, f_0; s)$  is a consequence of the functional equations of  $Z_\phi(s)$  and  $\Phi^{(i)}(f_\infty; \alpha, s)$ . In the present case, the functional equation of  $\Phi^{(i)}(f_\infty; \alpha, s)$  is a quite simple special case of [S6, Theorem 3.1]. The functional equation of  $Z_\phi(s)$  will be proved in §3. Combining these results, one can prove the analytic continuation and the functional equation of  $\zeta^{(i)}(\alpha, f_0; s)$ .

## 2.2 Epstein zeta functions with spherical functions

In compact case, a general theory of the zeta functions  $\zeta_i^{(i)}(\phi, f_0; s)$  have been developed in [S6]. Therefore, we give here only a simplest example of zeta functions of compact case, namely, the Epstein zeta function.

Let  $\mathrm{SO}(n) (n \geq 3)$  be the special orthogonal group for the quadratic form  $q(v) = v_1^2 + \dots + v_n^2$  and  $\rho_0$  the vector representation of  $\mathrm{SO}(n)$  on the  $n$ -dimensional vector space  $V$ . We identify  $V$  with the vector space of column vectors with  $n$  entries. Put  $\mathbf{G} = \mathrm{GL}(1) \times \mathrm{SO}(n)$  and let  $\rho$  be the rational representation of  $\mathbf{G}$  on  $V$  defined by

$$\rho(t, k)v = t\rho_0(k)v \quad (t \in \mathrm{GL}(1), k \in \mathrm{SO}(n), v \in V).$$

Then  $(\mathbf{G}, \rho, V)$  is a regular prehomogeneous vector space. We consider the natural  $\mathbb{Q}$ -structure on  $(\mathbf{G}, \rho, V)$ . We put  $\mathbf{L} = \mathbf{G}$ ,  $\mathbf{U} = \{1\}$ , and  $\Gamma = \{1\}$ . Then  $\mathbf{L}_0 = \mathrm{SO}(n)$ .

Since  $V_{\mathbf{R}} - S_{\mathbf{R}} = V_{\mathbf{R}} - \{0\}$ , we have  $v = 1$  and

$$X_1 = \mathrm{SO}(n)/\mathrm{SO}(n-1) \simeq \{v \in V_{\mathbf{R}} \mid q(v) = 1\}.$$

Here we identify the Lie group  $\mathrm{SO}(n-1)$  with the isotropy subgroup of  $\mathrm{SO}(n)$  at  $x_0 = (0, \dots, 0, 1)$ . For  $v \in V_{\mathbf{R}} - S_{\mathbf{R}}$ , the element  $\bar{v}$  in  $X_1$  is given by  $\bar{v} = q(v)^{-1/2}v$ .

Let  $\pi$  be an irreducible unitary representation of  $\mathrm{SO}(n) = \mathrm{SO}(n)_{\mathbf{R}}$  and  $\mathcal{A}(\mathrm{SO}(n), \pi)$  the space of functions  $\phi$  on  $\mathrm{SO}(n)$  with values in  $W_\pi$  satisfying  $\phi(kh) = \pi(k)\phi(h)$  ( $h, k \in \mathrm{SO}(n)$ ). Then the space  $\mathcal{A}(\mathrm{SO}(n), \pi)$  is isomorphic to  $W_\pi$  by the mapping  $\phi \mapsto \phi(1)$ . Since  $\mathcal{Z}(\mathrm{SO}(n))$  acts on  $\mathcal{A}(\mathrm{SO}(n), \pi)$  as scalar multiplication, we need not specify the infinitesimal character.

Recall that  $\pi$  is said to be of class one (with respect to  $\mathrm{SO}(n-1)$ ) if there exists a non-zero vector in  $W_\pi$  fixed under  $\mathrm{SO}(n-1)$ . It is not hard to see that  $Z_\phi(s)$  vanishes unless  $\pi$  is of class one (cf. [S4, Lemma 2.1]). Therefore, in the following, we assume that  $\pi$  is of class one. Take a non-zero vector  $w_0 \in W_\pi$  with norm 1 fixed under  $\mathrm{SO}(n-1)$ . By the irreducibility of  $\pi$ , such a  $w_0$  is unique up to a constant factor with absolute value 1. Then the space  $\mathcal{E}(X_1, \pi)$  of functions  $\psi$  on  $X_1$  satisfying  $\psi(k\bar{x}) = \pi(k)\psi(\bar{x})$  is 1-dimensional and is spanned by the function  $\psi_0(k\bar{x}_0) = \pi(k)w_0$ . Hence, for any  $\phi \in \mathcal{A}(\mathrm{SO}(n), \pi)$  and any  $x \in V_{\mathbb{Q}} - \{0\}$ , there exists a constant  $c(\phi, x)$  such that  $\mathcal{M}_x^{(1)}\phi(\bar{y}) = c(\phi, x)\psi_0(\bar{y})$ . It is obvious that  $c(\phi, x) = \langle \mathcal{M}_x^{(1)}\phi(\bar{x}_0), w_0 \rangle$ . On

the other hand  $\mathcal{M}_x^{(1)}\phi(\bar{x}_0) = \langle \phi(1), \psi_0(\bar{x}) \rangle w_0$ . Hence we have

$$c(\phi, x) = \langle \phi(1), \psi_0(\bar{x}) \rangle \text{ and } \mathcal{M}_x^{(1)}\phi(\bar{y}) = \langle \phi(1), \psi_0(\bar{x}) \rangle \psi_0(\bar{y}).$$

The zeta function and the local zeta function are defined as follows:

$$\zeta(\phi, f_0; s) = \sum_{x \in \mathbf{V}_Q - \{0\}} \frac{\mu(x) f_0(x) \langle \phi(1), \psi_0(\bar{x}) \rangle}{q(x)^s},$$

$$\Phi(f_\infty; \pi, s) = \int_{\mathbf{V}_R - \{0\}} q(y)^{s - (n/2)} \psi_0(\bar{y}) f_\infty(y) dy.$$

Moreover the integral representation of the zeta function given in Proposition 1.7 reads

$$\int_0^\infty t^{2s-1} dt \int_{SO(n)} \phi(k) \sum_{\mathbf{v}_Q \in S_Q} f_0(x) f_\infty(t\rho_0(k)x) dk$$

$$= \zeta(\phi, f_0; s) \Phi(f_\infty; \pi, s).$$

It is known that any representation  $\pi$  of class one can be realized on the space  $H_d$  of harmonic polynomials of homogeneous degree  $d$ ;  $d$  is determined by  $\pi$ . For  $w \in W_\pi$ , put

$$P_w(v) = d(\pi)^{1/2} q(v)^{d/2} \langle w, \psi_0(\bar{v}) \rangle,$$

where  $d(\pi) = \dim W_\pi$ . Then the mapping  $w \mapsto P_w$  defines a linear isomorphism of  $W_\pi$  onto  $H_d$ . Thus we have

$$\zeta(\phi, f_0; s) = d(\pi)^{-1/2} \sum_{x \in \mathbf{V}_Q - \{0\}} \frac{\mu(x) f_0(x) P_{\phi(1)}(x)}{q(x)^{s + (d/2)}}.$$

The right hand side of the identity above is the Epstein zeta function with spherical function (cf. [E] and [Si]).

In [S4], we proved the local functional equation of  $\Phi(f_\infty; \pi, s)$  and the global functional equation of  $\zeta(\phi, f_0; s)$  (for  $f_0$  = the characteristic function of a lattice in  $\mathbf{V}_Q$ ) from the same point of view as in the present paper. A generalization of this example is found in [S7].

### 3. Functional equation of the zeta integral

3.1 We keep the notation in § 1 and assume the conditions (A-1), (A-2) and (A-3). It is not necessary in the present section to assume (A-4). Instead we assume that

(A-5)  $(G, \rho, \mathbf{V})$  is decomposed over  $\mathbb{Q}$  into direct product as

$$(G, \rho, \mathbf{V}) = (G, \rho_1 \oplus \rho_2, \mathbf{E} \oplus \mathbf{F})$$

and the invariant subspace  $\mathbf{F}$  is a regular subspace.

For the definition and elementary properties of regular subspaces, we refer to [S1, § 2]. Note that, in [S1], we introduced the notion of  $k$ -regularity, where  $k$  is the field

of definition. However the  $\bar{k}$ -regularity implies the  $k$ -regularity (cf. [S6, §2.1]). Hence in the assumption (A-5),  $F$  is necessarily a  $\mathbb{Q}$ -regular subspace.

Let  $F^*$  be the vector space dual to  $F$  and  $\rho_2^*$  the rational representation of  $G$  on  $F^*$  contragredient to  $\rho_2$ . Set  $(G, \rho^*, V^*) = (G, \rho_1 \oplus \rho_2^*, E \oplus F^*)$ . The assumption (A-5) implies that  $(G, \rho^*, V^*)$  is also a prehomogeneous vector space defined over  $\mathbb{Q}$  and  $F^*$  is its regular subspace. By Lemma 2.4 in [S1], the assumption (A-1) holds also for  $(G, \rho^*, V^*)$ . Let  $S^*$  be the singular set of  $(G, \rho^*, V^*)$ . Let  $P_1^*, \dots, P_n^*$  be the basic relative invariants of  $(G, \rho^*, V^*)$  over  $\mathbb{Q}$ . Note that the number of basic relative invariants of  $(G, \rho^*, V^*)$  is equal to  $n$ , the number of basic relative invariants of  $(G, \rho, V)$ . Let  $\chi_i^*$  be the  $\mathbb{Q}$ -rational character of  $G$  corresponding to  $P_i^*$ :

$$P_i^*(\rho^*(g)x^*) = \chi_i^*(g)P_i^*(x^*) \quad (g \in G, x^* \in V^*).$$

Let  $X_{\rho^*}(G)_{\mathbb{Q}}$  be the subgroup of  $X(G)_{\mathbb{Q}}$  generated by  $\chi_1^*, \dots, \chi_n^*$ . Since  $X_{\rho}(G)_{\mathbb{Q}} = X_{\rho^*}(G)_{\mathbb{Q}}$ , there exists an  $n$  by  $n$  unimodular matrix  $U = (u_{ij})_{i,j=1}^n$  such that

$$\chi_i = \prod_{j=1}^n \chi_j^{u_{ij}} \quad (1 \leq i \leq n). \quad (3.1)$$

Let  $\lambda = (\lambda_1, \dots, \lambda_n)$  be an  $n$ -tuple of half-integers such that

$$(\det \rho_2(g))^2 = \prod_{i=1}^n \chi_i(g)^{2\lambda_i} \quad (3.2)$$

(for the existence of  $\lambda$ , see [S1, Lemma 2.5]).

Let  $\phi: L_0^+/\Gamma_L \rightarrow W$  be the same as in §1.4. Then, as in (1.2), we can define the zeta integral attached to  $\phi$  also for  $(G, \rho^*, V^*)$ :

$$\begin{aligned} Z_{\phi}^*(s)(f_{\infty}^* \otimes f_0^*) &= Z_{\phi}^*(s_1, \dots, s_n)(f_{\infty}^* \otimes f_0^*) \\ &= \frac{1}{v(\Gamma)} \int_{T^+} \prod_{i=1}^n \chi_i^*(t)^{s_i} \Delta(t) dt \int_{G_0^*/\Gamma} \phi(h) \\ &\quad \times \sum_{x^* \in V_0^* - S_0^*} f_0^*(x^*) f_{\infty}^*(\rho^*(thu)x^*) dh du \\ &\quad (f_{\infty}^* \otimes f_0^* \in \mathcal{S}(V_{\mathbb{R}}^*) \otimes \mathcal{S}(V_{\mathbb{Q}}^*)). \end{aligned}$$

Now let us recall the Poisson summation formula for functions in  $\mathcal{S}(V_{\mathbb{R}}) \otimes \mathcal{S}(V_{\mathbb{Q}})$ . For  $f_0 \in \mathcal{S}(V_{\mathbb{Q}})$  and  $x_2^* \in F_{\mathbb{Q}}$ , take a lattice  $\mathcal{L}$  in  $F_{\mathbb{Q}}$  such that the value of  $f_0(x_1, x_2)$  ( $x_1 \in E_{\mathbb{Q}}, x_2 \in F_{\mathbb{Q}}$ ) is determined by the coset of  $x_2$  modulo  $\mathcal{L}$  and  $x_2^*$  is contained in the dual lattice

$$\mathcal{L}^* = \{x_2^* \in F_{\mathbb{Q}}^* | \langle x_2^*, \mathcal{L} \rangle \subset \mathbb{Z}\}.$$

Put

$$\hat{f}_0(x_1, x_2^*) = v(\mathcal{L})^{-1} \sum_{x_2 \in F_{\mathbb{Q}}/\mathcal{L}} f_0(x_1, x_2) \exp[2\pi i \langle x_2, x_2^* \rangle],$$

where  $v(\mathcal{L}) = \int dx$ . Then  $\hat{f}_0(x_1, x_2^*)$  is independent of the choice of  $\mathcal{L}$  and defines

a function in  $\mathcal{S}(\mathbf{V}_Q^*)$ . The function  $f_0$  is called the partial Fourier transform of  $f_0$  with respect to  $\mathbf{F}$ .

We define the partial Fourier transform  $\hat{f}_\infty \in \mathcal{S}(\mathbf{V}_R^*)$  of  $f_\infty \in \mathcal{S}(\mathbf{V}_R)$  with respect to  $\mathbf{F}$  by setting

$$\hat{f}_\infty(x_1, x_2^*) = \int_{\mathbf{F}_R} f_\infty(x_1, x_2) \exp(-2\pi i \langle x_2, x_2^* \rangle) dx_2.$$

Then the partial Fourier transforms

$$\wedge: \mathcal{S}(\mathbf{V}_Q) \rightarrow \mathcal{S}(\mathbf{V}_Q^*) \text{ and } \wedge: \mathcal{S}(\mathbf{V}_R) \rightarrow \mathcal{S}(\mathbf{V}_R^*)$$

are linear isomorphisms and the following Poisson summation formula holds:

$$\begin{aligned} & \sum_{(x_1, x_2) \in \mathbf{V}_Q} f_0(x_1, x_2) f_\infty(\rho(g)(x_1, x_2)) \\ &= \det \rho_2(g)^{-1} \sum_{(x_1, x_2^*) \in \mathbf{V}_Q^*} \hat{f}_0(x_1, x_2^*) \hat{f}_\infty(\rho^*(g)(x_1, x_2^*)) \end{aligned} \quad (3.3)$$

$$(f_\infty \otimes f_0 \in \mathcal{S}(\mathbf{V}_R) \otimes \mathcal{S}(\mathbf{V}_Q), g \in G^+).$$

Let  $B$  (resp.  $B^*$ ) be the domain in  $\mathbb{C}^n$  on which  $Z_\phi(s)(f_\infty \otimes f_0)$  (resp.  $Z_\phi^*(s)(\hat{f}_\infty \otimes \hat{f}_0)$ ) converges absolutely. Denote by  $D$  (resp.  $D^*$ ) be the convex hull of  $(B^* U^{-1} + \lambda) \cup B$  (resp.  $(B - \lambda)U \cup B^*$ ) in  $\mathbb{C}^n$ . Then it is clear that  $(D - \lambda)U = D^*$ .

### PROPOSITION 3.1

Let  $f_\infty \in \mathcal{S}(\mathbf{V}_R)$  be a function satisfying that  $f_\infty$  and  $\hat{f}_\infty$  vanish identically on  $\mathbf{S}_R$  and  $\mathbf{S}_R^*$ , respectively. Then  $Z_\phi(s)(f_\infty \otimes f_0)$  and  $Z_\phi^*(s)(\hat{f}_\infty \otimes \hat{f}_0)$  have analytic continuations to holomorphic functions on  $D$  and  $D^*$ , respectively, and satisfy the functional equation

$$Z_\phi^*((s - \lambda)U)(\hat{f}_\infty \otimes \hat{f}_0) = Z_\phi(s)(f_\infty \otimes f_0) \quad (s \in D).$$

The proof of Proposition 3.1, which is based on (3.3), is quite similar to that of [S1, Lemma 6.1] and we do not repeat it here.

For later convenience, we recall the construction of functions  $f_\infty$  satisfying the assumption in Proposition 3.1. Let  $r = n - \text{rank } X_{\rho_1}(\mathbf{G})_Q$ . Then, among the basic relative invariants  $P_1, \dots, P_n$  of  $(\mathbf{G}, \rho, \mathbf{V})$  (resp.  $P_1^*, \dots, P_n^*$  of  $(\mathbf{G}, \rho^*, \mathbf{V}^*)$ ) over  $\mathbf{Q}$ , there exist precisely  $n - r$  relative invariants which are constant as functions of  $x_2$  on  $\mathbf{F}$  (resp.  $x_2^*$  on  $\mathbf{F}^*$ ). Hence we may assume that

$$P_i(x_1, x_2) = P_i^*(x_1, x_2^*) = P_i(x_1) \quad (i = r + 1, \dots, n).$$

These  $P_i(x_1)$  ( $r + 1 \leq i \leq n$ ) are the basic relative invariants of  $(\mathbf{G}, \rho_1, \mathbf{E})$  over  $\mathbf{Q}$ . We put

$$P_F(x_1, x_2) = \prod_{i=1}^r P_i(x_1, x_2) \text{ and } P_F^*(x_1, x_2^*) = \prod_{i=1}^r P_i^*(x_1, x_2^*).$$

**Lemma 3.2.** ([S1, Lemma 6.2]) (i) For an  $f'^* \in C_0^\infty(\mathbf{V}_R^* - \mathbf{S}_R^*)$ , put

$$f_\infty = P_F(x_1, x_2) \cdot \hat{f}'^*(x_1, x_2).$$

Then  $f_\infty$  and  $\hat{f}_\infty$  vanish on  $S_\mathbb{R}$  and  $S_\mathbb{R}^*$ , respectively.

(ii) For an  $f'_\infty \in C_0^\infty(V_\mathbb{R} - S_\mathbb{R})$ , put

$$f_\infty = P_F^* \left( x_1, \frac{\partial}{\partial x_2} \right) f'_\infty(x_1, x_2).$$

Then  $f_\infty$  and  $\hat{f}_\infty$  vanish on  $S_\mathbb{R}$  and  $S_\mathbb{R}^*$ , respectively.

#### 4. Prehomogeneous vector spaces with symmetric structure

4.1 In this section, we keep the notation in § 1 and assume the conditions (A-1) and (A-2). As in § 1.4, let  $p: G_0 \rightarrow L_0$  be the canonical surjection and put  $L_{(x)} = p(G_x \cap G_0)$  for  $x \in V - S$ .

We say that the semi-direct product decomposition  $G = LU$  determines a *symmetric structure* on  $(G, \rho, V)$  over  $\mathbb{Q}$  if, for any  $x \in V_\mathbb{Q} - S_\mathbb{Q}$ , there exists an involution (= an automorphism of order 2)  $\sigma: L_0 \rightarrow L_0$  defined over  $\mathbb{Q}$  such that

$$L_0^\sigma := \{h \in L_0 \mid \sigma(h) = h\} \supset L_{(x)} \supset (L_0^\sigma)^0. \quad (4.1)$$

Then, for any  $x \in V_\mathbb{R} - S_\mathbb{R}$ , there exists an involution  $\sigma$  of  $L_0$  defined over  $\mathbb{R}$  satisfying (4.1). The involution  $\sigma$  induces an involution of  $L_0^+$ , which we denote also by  $\sigma$ , satisfying

$$(L_0^+)^{\sigma} \supset L_0^+ \cap L_{(x)} = L_{(x)}^+ \supset ((L_0^+)^{\sigma})^0.$$

Therefore the homogeneous spaces  $X_i (1 \leq i \leq v)$  defined in § 1.5 are reductive symmetric spaces (symmetric case in § 1.6) and the construction of zeta functions given in § 1 can be applied to  $(G, \rho, V)$  with symmetric structure.

**Lemma 4.1** Suppose that  $(G, \rho, V)$  satisfies the condition (A-5) in § 2, namely,  $V$  contains a regular subspace  $F$ . Then the decomposition  $G = LU$  determines a symmetric structure also on  $(G, \rho^*, V^*)$ , the prehomogeneous vector space dual to  $(G, \rho, V)$  with respect to  $F$ .

*Proof.* By (A-5), one can find a relative invariant  $P$  of  $(G, \rho, V)$  with coefficients in  $\mathbb{Q}$  for which the rational mapping  $\phi_P: V - S \rightarrow V^*$  defined by

$$\phi_P(x_1, x_2) = (x_1, \text{grad}_{x_2}(\log P(x_1, x_2)))$$

gives rise to a  $G$ -equivariant biregular mapping of  $V - S$  onto  $V^* - S^*$  defined over  $\mathbb{Q}$ . For  $x \in V_\mathbb{Q} - S_\mathbb{Q}$ , put  $x^* = \phi_P(x) \in V_\mathbb{Q}^* - S_\mathbb{Q}^*$ . Then we have  $G_x = G_{x^*}$  and  $L_{(x)} = L_{(x^*)}$  (cf. [S1, Lemma 2.4]). Now the assertion is obvious. ■

**Examples.** (1) Let  $(G, \rho, V) = (GL(n), \rho, \text{Sym}(n))$ , where  $\text{Sym}(n)$  is the space of symmetric matrices of size  $n$  and the representation  $\rho$  is given by  $\rho(g)x = gx^t g (g \in GL(n), x \in \text{Sym}(n))$ . Then the trivial decomposition  $G = LU$  with  $L = GL(n)$  and  $U = \{1\}$  determines a symmetric structure. In fact, we have  $L_0 = SL(n)$  and, if we define  $\sigma(h) = x^t h^{-1} x^{-1} (h \in L_0)$ , then  $L_0^\sigma = L_{(x)} \simeq SO(n)$  for  $x \in V_\mathbb{Q} - S_\mathbb{Q}$ . The corresponding symmetric space is  $SL(n)/SO(n)$ .

(2) More generally, if  $(G, \rho, V)$  is an irreducible regular prehomogeneous vector space of commutative parabolic type, then the trivial decomposition  $G = LU$  with  $L = G$  and  $U = \{1\}$  determines a symmetric structure (cf. [Ru] and [BR]).

(3) Let  $(G, \rho, V) = (GL(1) \times SO(m) \times SL(n), \rho, M(m, n))$  ( $m \geq n \geq 1$ ), where the representation  $\rho$  is given by

$$\rho(t, k, g)x = tkxg^{-1} \quad (t \in GL(1), g \in SL(n), k \in SO(m), x \in M(m, n)).$$

Then one can define two symmetric structures on  $(G, \rho, V)$ . In fact, the following two decompositions

$$G = L \times U, L = GL(1) \times SO(m), U = SL(n)$$

and

$$G = L \times U, L = GL(1) \times SL(n), U = SO(m)$$

determine different symmetric structures. The corresponding symmetric spaces are

$$SO(m)/(SO(n) \times SO(m-n)) \quad \text{in the former case,}$$

$$SL(n)/SO(n) \quad \text{in the latter case.}$$

**4.2** Let  $P_L$  be a parabolic subgroup of  $L$  and put  $P = P_L U$ . We denote the restriction of the representation  $\rho$  to  $P$  by the same symbol  $\rho$ . We do not assume that  $P_L$  is defined over  $\mathbb{Q}$ . In fact, in § 5, we need to consider a parabolic subgroup defined over  $\mathbb{R}$ .

*Lemma 4.2.* Suppose that  $G = LU$  determines a symmetric structure of  $(G, \rho, V)$ . Then

- (1)  $(P, \rho, V)$  is also a prehomogeneous vector space.
- (2) If  $(G, \rho, V)$  is regular, so is  $(P, \rho, V)$ .

*Proof.* Let  $V_1 = \{x \in V \mid P_1(x) = \cdots = P_n(x) = 1\}$ . Then  $V_1$  is a single  $\rho(G_0)$ -orbit (cf. [S6, Lemma 1.1]). Fix a point  $x_0 \in V_1$ . The mapping  $\beta: V_1 \rightarrow L_0/L_{(x_0)}$  defined by  $\beta(\rho(hu)x_0) = h \cdot L_{(x_0)}$  ( $h \in L_0, u \in U$ ) is clearly  $L_0$ -equivariant. Since  $L_0/L_{(x_0)}$  is a symmetric space and  $P_{L_0} = P_L \cap L_0$  is a parabolic subgroup of  $L_0$ , there exists a Zariski-open  $P_{L_0}$ -orbit  $\Omega_0$  in  $L_0/L_{(x_0)}$  (see [V, § 1]). Then  $\Omega = \rho(T)\beta^{-1}(\Omega_0)$  is a Zariski-open  $P$ -orbit in  $V$ . Hence  $(P, \rho, V)$  is a prehomogeneous vector space. The second assertion is obvious. ■

We denote by  $S_P$  the singular set of  $(P, \rho, V)$ . It is obvious that  $S_P \supset S$ . Recall that the parabolic subgroup  $P_{L_0} = P \cap L_0$  of  $L_0$  is called  $\sigma$ -anisotropic for an involution  $\sigma$  of  $L_0$  if  $P_{L_0} \cap \sigma(P_{L_0})$  is a Levi subgroup of  $P_{L_0}$  (cf. [V, § 1]).

*Lemma 4.3.* Suppose that  $G = LU$  determines a symmetric structure of  $(G, \rho, V)$  and  $P_{L_0} = P_L \cap L_0$  is  $\sigma$ -anisotropic for the involution  $\sigma$  corresponding to some  $x_0 \in V - S$ . Then

- (1) the point  $x_0$  is in  $V - S_P$ .
- (2) For  $x \in V - S_P$ , the isotropy subgroup  $P_x = \{p \in P \mid \rho(p)x = x\}$  is (not necessarily connected) reductive.
- (3) The singular set  $S_P$  is a hypersurface.



*Proof.* We use the notation in the proof of Lemma 4.2. By replacing  $x_0$  by  $\rho(t)x_0 (t \in T)$  if necessary, we may assume that  $x_0 \in V_1$ . By [V, Theorem 1] and the assumption that  $P_{L_0}$  is  $\sigma$ -anisotropic, we see that  $\beta(x_0)$  is in  $\Omega_0$ . This implies the first assertion. To prove the second assertion, it is sufficient to consider the case where  $x = x_0$ . Since the identity component of  $P_{x_0}$  coincides with that of  $(P_{L_0} \cdot U)_{x_0} = P_{x_0} \cap (P_{L_0} \cdot U)$ , we prove that  $(P_{L_0} \cdot U)_{x_0}$  is reductive. By (A-1),  $G_{x_0}$  is reductive; hence its normal subgroup  $U_{x_0} = U \cap G_{x_0}$  is reductive. Put  $L'_0 = P_{L_0} \cap \sigma(P_{L_0})$ . By the assumption,  $L'_0$  is a Levi subgroup of  $P_{L_0}$ . The identity component of  $P_{L_0} \cap L_{(x_0)}$  coincides with that of  $(L'_0)^\sigma$ . Hence the group  $P_{L_0} \cap L_{(x_0)}$  is reductive (cf. [V, § 1]). Since  $(P_{L_0} \cdot U)_{x_0}/U_{x_0} \simeq P_{L_0} \cap L_{(x_0)}$ , this proves the second part. The third assertion is an immediate consequence of the second. ■

*Remarks.* (1) The prehomogeneous vector space  $(P, \rho, V)$  does not necessarily satisfy the latter half of (A-1), even if it is defined over  $\mathbb{Q}$ .

(2) Under the assumption in Lemma 4.3, we put

$$L'_0 = P_{L_0} \cap \sigma(P_{L_0}), \quad L' = TL'_0, \quad U' = R_u(P_L)U.$$

As the following example illustrates, the semi-direct product decomposition  $P = L'U'$  determines a symmetric structure of  $(P, \rho, V)$ .

*Example.* (1) Let  $(G, \rho, V)$  be as in Example 1 and put

$$P = \left\{ g = \begin{pmatrix} g_1 & 0^{(p, n-p)} \\ g_3 & g_4 \end{pmatrix} \in G = \right. \\ \left. GL(m) \left| \begin{array}{l} g_1 \in GL(p), g_3 \in M(n-p, p), g_4 \in GL(n-p) \end{array} \right. \right\}.$$

Then  $P_{L_0} = P \cap SL(n)$  is  $\sigma$ -anisotropic, where  $\sigma$  is the involution given by  $\sigma(g) = {}^t g^{-1}$ . We have

$$L' = \left\{ \begin{pmatrix} g_1 & 0 \\ 0 & g_4 \end{pmatrix} \left| \begin{array}{l} g_1 \in GL(p), g_4 \in GL(n-p) \end{array} \right. \right\} \cong GL(p) \times GL(n-p) \\ U' = \left\{ \begin{pmatrix} 1_p & 0 \\ g_3 & 1_{n-p} \end{pmatrix} \left| \begin{array}{l} g_3 \in M(n-p, p) \end{array} \right. \right\}.$$

The basic relative invariants of  $(P, \rho, V)$  are given by

$$P_1(x) = \det(x_1), \quad \chi_1(g) = \det(g_1)^2, \\ P_2(x) = \det(x), \quad \chi_2(g) = \det(g)^2,$$

where  $x_1$  is the upper left  $p$  by  $p$  block of  $x$ . We have

$$S_P = \{x \in V \mid P_1(x) = 0\} \cup \{x \in V \mid P_2(x) = 0\}.$$

For  $x_0 = 1_n \in V - S_P$ , the corresponding symmetric space is

$$(SL(p) \times SL(n-p)) / (SO(p) \times SO(n-p)) \\ = (SL(p)/SO(p)) \times (SL(n-p)/SO(n-p)).$$

4.3 Let the assumption be as in Lemma 4.3. Take a field  $k$  such that  $\mathbf{P}_{L_0}$  and the involution  $\sigma$  are defined over  $k$ , and  $x_0 \in V_k - S_k$ . We examine the group  $X_\rho(\mathbf{P})_k$  of  $k$ -rational characters corresponding to relative invariants of  $(\mathbf{P}, \rho, V)$ .

For simplicity, we assume that

(A-6) the basic relative invariants  $P_1, \dots, P_n$  of  $(\mathbf{G}, \rho, V)$  over  $\mathbb{Q}$  are absolutely irreducible.

This is equivalent to the condition

$$X_\rho(\mathbf{G})_{\mathbb{Q}} = X_\rho(\mathbf{G})_{\mathbb{C}}.$$

Let  $\mathbf{T}'_0$  be the identity component of the center of  $\mathbf{L}'_0$ . The central torus  $\mathbf{T}'_0$  is  $\sigma$ -stable. Hence we get a separable isogeny  $\mathbf{T}'_{0,+} \times \mathbf{T}'_{0,-} \rightarrow \mathbf{T}'_0$ , where

$$\mathbf{T}'_{0,+} = \{t \in \mathbf{T}'_0 \mid \sigma(t) = t\}^0 \text{ and } \mathbf{T}'_{0,-} = \{t \in \mathbf{T}'_0 \mid \sigma(t) = t^{-1}\}^0.$$

We consider the following commutative diagram of the natural mappings:

$$\begin{array}{ccc} X(\mathbf{P})_k & \hookrightarrow & X(\mathbf{P})_k \oplus X(\mathbf{T}'_0)_k \\ \uparrow & & \downarrow \text{restriction to } \mathbf{T} \times \mathbf{T}'_{0,-} \\ X_\rho(\mathbf{P})_k & \xrightarrow{\xi} & X(\mathbf{T})_k \oplus X(\mathbf{T}'_{0,-})_k. \end{array}$$

Here note that  $X(\mathbf{T})_k = X(\mathbf{T})_{\mathbb{Q}}$ , since  $\mathbf{T}$  is a  $\mathbb{Q}$ -split torus.

*Lemma 4.4. The homomorphism  $\xi: X_\rho(\mathbf{P})_k \rightarrow X(\mathbf{T})_k \oplus X(\mathbf{T}'_{0,-})_k$  is injective and of finite cokernel.*

*Proof.* Any character  $\chi$  in  $X_\rho(\mathbf{P})_k$  is trivial on  $\mathbf{P}_{x_0}\mathbf{U}$ . The group  $\mathbf{P}_{x_0}\mathbf{U}$  contains  $(\mathbf{P}_{x_0} \cap \mathbf{L}_{(x_0)})\mathbf{U}$  and hence  $((\mathbf{L}'_0)^\sigma)^0\mathbf{U}$ . Since  $\mathbf{T}'_{0,+}$  is a subgroup of  $((\mathbf{L}'_0)^\sigma)^0$ ,  $\chi$  is trivial on  $\mathbf{T}'_{0,+}$ . This implies that  $\xi$  is injective. As we have already seen in § 1.2,  $\xi(X_\rho(\mathbf{G})_k) (= \xi(X_\rho(\mathbf{G})_{\mathbb{Q}}))$  by (A-6) is of finite index in  $X(\mathbf{T})_k (= X(\mathbf{T})_{\mathbb{Q}})$ . Let  $\chi$  be a  $k$ -rational character of  $\mathbf{T}'_{0,-}$ . Then, for some integer  $e_1$ ,  $\chi^{e_1}$  can be extended to a  $k$ -rational character of  $\mathbf{P}$  such that  $\ker \chi^{e_1}$  contains  $\mathbf{T}\mathbf{T}'_{0,+}D(\mathbf{L}'_0)\mathbf{U}'$ , where  $D(\mathbf{L}'_0)$  is the derived group of  $\mathbf{L}'_0$ . Since  $(\mathbf{P}_x)^0$  is contained in  $\mathbf{T}\mathbf{T}'_{0,+}D(\mathbf{L}'_0)\mathbf{U}'$ , there exists an integer  $e$  such that  $\chi^e$  is trivial on  $\mathbf{P}_x$ . This implies that  $\chi^e \in X_\rho(\mathbf{P})_k$ . Therefore  $\xi(X_\rho(\mathbf{P})_k)$  is of finite index in  $X(\mathbf{T})_k \oplus X(\mathbf{T}'_{0,-})_k$ . ■

Let  $P_1, \dots, P_n, P_{n+1}, \dots, P_{n+l}$  be the basic relative invariants of  $(\mathbf{P}, \rho, V)$  over  $k$ , where  $P_1, \dots, P_n$  are the basic relative invariants of  $(\mathbf{G}, \rho, V)$ . We have  $l = \text{rank } X(\mathbf{T}'_{0,-})_k$  by Lemma 4.4.

Let  $\chi_{n+1}, \dots, \chi_{n+l}$  be the  $k$ -rational characters corresponding to  $P_{n+1}, \dots, P_{n+l}$ , respectively. Take a positive integer  $e$  such that  $(\chi_i^e|_{\mathbf{T}})$  ( $n+1 \leq i \leq n+l$ ) are in  $\xi(X_\rho(\mathbf{G})_k)$ . Then one can find  $m_{ij} \in e^{-1}\mathbb{Z}$  ( $1 \leq i \leq n$ ,  $1 \leq j \leq l$ ) such that

$$\chi_{n+j}^e \left/ \prod_{i=1}^n \chi_i^{e m_{ij}} \right. \equiv 1 \quad \text{on } \mathbf{T}. \quad (4.2)$$

These  $m_{ij}$  will play a role in the algebraic construction of the Poisson kernel in § 5.

Let  $(G, \rho, V)$  be a prehomogeneous vector space with symmetric structure  $G = L \cdot U$  satisfying the assumptions (A-1), (A-3), (A-5) and (A-6). The assumption (A-2) is automatically satisfied. In this section we prove the functional equation satisfied by zeta functions attached to automorphic forms under the following assumption:

(A-7)  $L_0$  is semisimple and the symmetric spaces  $X_i = L_0^+ / L_{(x_i)}^+$  ( $1 \leq i \leq v$ ) are  $K_\varepsilon$ -spaces in the sense of [OS].

**5.1** Let  $K$  be a maximal compact subgroup of  $L_0^+$  and  $\theta$  the corresponding Cartan involution. Let  $P_0$  be a minimal parabolic subgroup of  $L_0^+$  with Langlands decomposition  $P_0 = MAN$  with respect to  $\theta$ . Denote by  $\mathfrak{L}_0$ ,  $\mathfrak{m}$  and  $\mathfrak{a}$  the Lie algebras of  $L_0^+$ ,  $M$  and  $A$ , respectively. Let  $\Sigma(\subset \mathfrak{a}^*)$  be the set of restricted roots and  $\Sigma^+$  the set of positive restricted roots corresponding to  $P_0$ . Put  $\Omega_0^\alpha = \{X \in \mathfrak{L}_0 \mid [H, X] = \alpha(H)X, H \in \mathfrak{a}\}$  for  $\alpha \in \Sigma$ .

Following [OS], we call a mapping  $\varepsilon: \Sigma \rightarrow \{\pm 1\}$  a signature of roots if it satisfies the condition

$$\begin{aligned} \varepsilon(\alpha) &= \varepsilon(-\alpha) & (\alpha \in \Sigma), \\ \varepsilon(\alpha + \beta) &= \varepsilon(\alpha)\varepsilon(\beta) & \text{if } \alpha, \beta \in \Sigma \text{ and } \alpha + \beta \in \Sigma. \end{aligned}$$

For a signature of roots  $\varepsilon$ , define an involution  $\theta_\varepsilon$  of  $\mathfrak{L}_0$  by

$$\begin{aligned} \theta_\varepsilon(X) &= \varepsilon(-\alpha)\theta(X) & X \in \Omega_0^\alpha, \alpha \in \Sigma, \\ \theta_\varepsilon(X) &= \theta(X) & X \in \mathfrak{m} + \mathfrak{a}. \end{aligned}$$

Then a precise formulation of the condition (A-7) is as follows:

(A-7)' for each  $i = 1, \dots, v$ , there exists a representative  $x_i \in V_i$  and a signature of roots  $\varepsilon_i$  such that  $L_{(x_i)}^+ = M \cdot K_{\varepsilon_i}^0$ , where  $K_{\varepsilon_i}^0$  is the connected analytic subgroup of  $L_0^+$  with Lie algebra

$$\mathfrak{k}_{\varepsilon_i} = \{X \in \mathfrak{L}_0 \mid \theta_{\varepsilon_i}(X) = X\}.$$

In this case, one can apply the results in [OS] to the homogeneous spaces  $X_i = L_0^+ / L_{(x_i)}^+$ .

Let  $W = N_K(A) / Z_K(A)$  be the Weyl group. Note the  $M = Z_K(A)$ . Define a subgroup  $W^{(i)}$  of  $W$  by  $W^{(i)} = (L_{(x_i)}^+ \cap N_K(A)) / M$ . Put  $r_i = [W : W^{(i)}]$  and fix a complete system  $\{w_1^{(i)}, \dots, w_{r_i}^{(i)}\}$  of representatives of  $W / W^{(i)}$ . Then, by [OS, Proposition 1.10] (or by [Mat]), the set

$$\bigcup_{j=1}^{r_i} AN w_j^{(i)} L_{(x_i)}^+ = \bigcup_{j=1}^{r_i} P_0 w_j^{(i)} L_{(x_i)}^+ \quad (\text{disjoint union}) \quad (5.1)$$

is an open dense subset of  $L_0^+$ .

Let  $P_{L_0}$  be a minimal  $\mathbb{R}$ -parabolic subgroup of  $L_0$  such that  $P_{L_0, \mathbb{R}} \cap L_0^+ = P_0$ . The parabolic subgroup  $P_{L_0}$  is  $\theta_\varepsilon$ -anisotropic. Put

$$P = TP_{L_0}U \text{ and } P^+ = T^+P_0U^+.$$

Note that  $P^+$  is not necessarily connected. Using the notation in §4.3, we have  $T'_0 = T'_{0,-}$  and  $T'_{0,+} = \{1\}$ .

As in §4.2, let  $S_P$  be the singular set of  $(P, \rho, V)$ . Then it follows from (5.1) that the  $P^+$ -orbit decomposition of  $V_R - S_{P,R}$  is given by

$$V_R - S_{P,R} = \bigcup_{i=1}^v \bigcup_{j=1}^{r_i} V_{ij}, \quad V_{ij} = \rho(P^+)x_{ij}, \quad x_{ij} = \rho(w_j^{(i)})x_i.$$

Let  $P_1, \dots, P_n, P_{n+1}, \dots, P_{n+l}$  be the basic relative invariants of  $(P, \rho, V)$  over  $\mathbb{R}$ . As in §4.3,  $P_1, \dots, P_n$  are the basic relative invariants of  $(G, \rho, V)$ . We have  $l = \dim A$  in the present case. Let  $m_{ij}$  ( $1 \leq i \leq n, 1 \leq j \leq l$ ) be the rational numbers given by (4.2). Then the function

$$|p_j(x)| = |P_{n+j}(x)| \left/ \prod_{i=1}^n |P_i(x)|^{m_{ij}} \right. \quad (i \leq j \leq l)$$

on  $V_R - S_R$  satisfies that

$$|p_j(\rho(tmanu)x)| = \chi_{n+j}(a) |p_j(x)| \quad (t \in T^+, m \in M, a \in A, n \in N, u \in U^+, x \in V_R - S_R).$$

This implies that  $|p_j|$  defines a function  $|\bar{p}_j|$  on  $X_i$ :

$$\begin{array}{ccc} V_i & \xrightarrow{|p_j|} & \mathbb{R}_+^\times \\ & \searrow & \nearrow \\ & X_i & \end{array} \quad \begin{array}{c} \\ \\ |\bar{p}_j| \end{array}$$

By Lemma 4.4,  $\{\chi_{n+1}, \dots, \chi_{n+l}\}$  gives a basis of  $X(T'_0)_R \otimes \mathbb{C}$ . We can identify  $X(T'_0)_R \otimes \mathbb{C}$  with  $\mathfrak{a}_C^* = \mathfrak{a}^* \otimes_R \mathbb{C}$  by  $X(T'_0)_R \ni \chi \mapsto \log(\chi \circ \exp) \in \mathfrak{a}_C^*$ . For  $\lambda \in \mathfrak{a}_C^*$ , write  $\lambda = \sum_{j=1}^l \lambda_j \log(\chi_{n+j} \circ \exp)$  and put

$$|p(x)|^\lambda = \prod_{j=1}^l |p_j(x)|^{\lambda_j} \quad (x \in V_R - S_R)$$

and

$$|p(x)|_{ij}^\lambda = \begin{cases} |p(x)|^\lambda & \text{if } x \in V_{ij}, \\ 0 & \text{otherwise.} \end{cases}$$

The function  $|p(x)|_{ij}^\lambda$  is well-defined for  $\operatorname{Re}(\lambda_1), \dots, \operatorname{Re}(\lambda_l) > 0$  and we define  $|p(x)|_{ij}^\lambda$  for arbitrary  $\lambda \in \mathfrak{a}_C^*$  by analytic continuation. We denote by  $|\bar{p}(\bar{x})|_{ij}^\lambda$  the function on  $X_i$  induced by  $|p(x)|_{ij}^\lambda$ . Then  $|\bar{p}(\bar{x})|_{ij}^\lambda$  ( $\bar{x} = h\bar{x}_i, h \in L_0^+$ ) ( $1 \leq j \leq r_l$ ) coincide with the functions  $\exp\{-\lambda(H_{e_i}^{w_0}(h^{-1}))\}$  defined by [OS, (3.33)].

**5.2** Now we examine the space  $\mathcal{S}(X_i; \pi, \chi)$  of spherical functions of type  $(\pi, \chi)$  introduced in §1.5. Let  $\mathcal{D}(X_i) = \mathcal{D}(L_0^+ / L_{(x_i)}^+)$  be the algebra of  $L_0^+$ -invariant differential operators on  $X_i$ . Denote by  $\mathcal{Z}(X_i)$  the subring of  $\mathcal{D}(X_i)$  consisting of the restrictions  $\bar{D}$  of bi-invariant differential operators  $D$  in  $\mathcal{Z}(L_0^+)$ . It is known that  $\mathcal{Z}(X_i) = \mathcal{D}(X_i)$  if  $L_0^+$  is of classical type and  $\mathcal{D}(X_i)$  is a finite  $\mathcal{Z}(X_i)$ -module in general ([Hel1, § 7], [Hel4]).

Let

$$\gamma_i: \mathcal{D}(X_i) (\cong U(\mathfrak{Q}_0)^{L_{i,0}^+} / (U(\mathfrak{Q}_0)^{L_{i,0}^+} \cap U(\mathfrak{Q}_0)(\mathfrak{k}_{e_i}))) \xrightarrow{\cong} U(\mathfrak{a})^W$$

be the standard isomorphism of  $\mathcal{D}(X_i)$  onto the ring  $U(\mathfrak{a})^W$  of the Weyl group invariants (cf. [OS, § 2.3], [Hel3, Chap. II, § 4, § 5]). For  $\mu \in \mathfrak{a}_{\mathbb{C}}^*$ , we obtain an algebra homomorphism of  $U(\mathfrak{a})^W$  into  $\mathbb{C}$  by extending it to  $U(\mathfrak{a})^W$ , which we denote by the same symbol. Put

$$\chi_\mu := \mu \circ \gamma_i: \mathcal{D}(X_i) \rightarrow \mathbb{C}. \quad (5.2)$$

Let  $\pi$  be an irreducible unitary representation of  $K$  on a finite-dimensional Hilbert space  $W_\pi$  and  $\chi: \mathcal{Z}(L_0^+) \rightarrow \mathbb{C}$  an infinitesimal character. It is obvious that  $\mathcal{E}(X_i; \pi, \chi) = \{0\}$  unless  $\chi: \mathcal{Z}(L_0^+) \rightarrow \mathbb{C}$  factors through  $\mathcal{Z}(X_i) = \mathcal{Z}(L_0^+ / L_{(x_i)}^+)$ :

$$\begin{array}{ccc} \chi: \mathcal{Z}(L_0^+) & \longrightarrow & \mathbb{C} \\ & \searrow & \nearrow \\ & \mathcal{Z}(L_0^+ / L_{(x_i)}^+) & \end{array}$$

In the following, we assume that  $\mathcal{E}(X_i; \pi, \chi) \neq \{0\}$  and denote the character of  $\mathcal{Z}(X_i)$  induced by  $\chi$  also by the same symbol.

For  $\mu \in \mathfrak{a}_{\mathbb{C}}^*$ , put

$$\mathcal{E}_0(X_i; \pi, \chi_\mu) = \left\{ \psi: X_i \rightarrow W_\pi \left| \begin{array}{ll} \psi(k\bar{x}) = \pi(k)\psi(\bar{x}) & (k \in K, \bar{x} \in X_i) \\ D\psi = \chi_\mu(D)\psi & (D \in \mathcal{D}(X_i)) \end{array} \right. \right\}.$$

Since  $\mathcal{D}(X_i) \supset \mathcal{Z}(X_i)$ , we have

$$\mathcal{E}_0(X_i; \pi, \chi_\mu) \subseteq \mathcal{E}(X_i; \pi, \chi_\mu|_{\mathcal{Z}(X_i)}).$$

On the other hand, since  $\mathcal{D}(X_i)$  is a commutative algebra, the ring  $\mathcal{D}(X_i)$  acts on  $\mathcal{E}(X_i; \pi, \chi)$  ( $\chi \in \text{Hom}(\mathcal{Z}(L_0^+), \mathbb{C})$ ). We assume that the action of  $\mathcal{D}(X_i)$  on  $\mathcal{E}(X_i; \pi, \chi)$  is semisimple, namely,

(A-8) there exist a finite number of  $\mu_1, \dots, \mu_d \in \mathfrak{a}_{\mathbb{C}}^*$  such that

$$\mathcal{E}(X_i; \pi, \chi) = \bigoplus_{j=1}^d \mathcal{E}_0(X_i; \pi, \chi_{\mu_j}). \quad (5.3)$$

*Remark.* The assumption always holds unless the symmetric space  $X_i$  contains an irreducible symmetric space of type EVII or EIX ([Och]). Even in this exceptional case, the assumption holds for generic  $\chi$ . By [OS, Lemma 2.24],  $\mu_1, \dots, \mu_d$  do not depend on  $i = 1, \dots, v$ . If  $G$  is of classical type, then  $\mathcal{D}(X_i) = \mathcal{Z}(X_i)$  and  $d = 1$  for any  $\chi$ . In some exceptional cases, it may occur that  $d \geq 2$ ; however, for a generic  $\chi$ , we may take  $d = 1$  ([Hel4]).

Now we define  $\text{End}(W_\pi)$ -valued spherical functions  $\Psi_{i,j}^{\pi, \mu}(\bar{x})$  ( $\bar{x} \in X_i$ ) by analytic continuation (as a function of  $\mu$ ) of the integral

$$\Psi_{i,j}^{\pi, \mu}(\bar{x}) := \int_K |\bar{p}(k^{-1} \cdot \bar{x})|_{ij}^{\mu + \rho} \pi(k) dk \quad (\mu \in \mathfrak{a}_{\mathbb{C}}^*, 1 \leq i \leq v, 1 \leq j \leq r_i), \quad (5.4)$$

where  $\rho = \frac{1}{2} \sum_{\alpha \in \Sigma^+} \alpha$  and  $dk$  is the normalized Haar measure on  $K$ . (We have used  $\rho$  to denote a prehomogeneous representation of  $G$ . This double use of the symbol will not cause any confusion, since the meaning is clear from the context.) Then, using the Poisson integral representation of eigenfunctions on  $X_i$  of invariant differential operators ([OS, Theorem 5.1]), we immediately obtain the following proposition:

### PROPOSITION 5.1

If  $\mu \in \mathfrak{a}_\mathbb{C}^*$  satisfies  $2\langle \mu, \alpha \rangle / \langle \alpha, \alpha \rangle \notin \mathbb{Z}$  for all  $\alpha \in \Sigma$ , then  $\Psi_{i,j}^{\pi,\mu}(\bar{x})$  is holomorphic at  $\mu$  and the linear mapping

$$\begin{aligned} \mathcal{P}_{i,\mu} : (W_\pi^M)^{\oplus r_i} &\rightarrow \mathcal{E}_0(X_i; \pi, \chi_\mu) \\ (v_j)_{j=1}^{r_i} &\mapsto \sum_{j=1}^{r_i} \Psi_{i,j}^{\pi,\mu}(\bar{x}) v_j \end{aligned}$$

is an isomorphism, where

$$W_\pi^M = \{v \in W_\pi \mid \pi(m)v = v (m \in M)\}.$$

Thus we have constructed a basis of  $\mathcal{E}(X_i; \pi, \chi)$  for generic  $\mu \in \mathfrak{a}_\mathbb{C}^*$ .

Let  $\mu_1, \dots, \mu_d$  be the elements in  $\mathfrak{a}_\mathbb{C}^*$  appearing in the right hand side of the decomposition (5.3). We assume in the following that  $2\langle \mu_l, \alpha \rangle / \langle \alpha, \alpha \rangle \notin \mathbb{Z}$  for all  $\alpha \in \Sigma$  and  $1 \leq l \leq d$ . Then, for  $\phi \in \mathcal{A}(L_0^+ / \Gamma; \pi, \chi)$  and  $x \in V_Q \cap V_i$ , one can find elements  $v_{j,\mu_l}^{(i)}(\phi; x) \in W_\pi^M$  such that

$$\mathcal{M}_x^{(i)} \phi(\bar{y}) = \sum_{l=1}^d \sum_{j=1}^{r_l} \Psi_{i,j}^{\pi,\mu_l}(\bar{y}) v_{j,\mu_l}^{(i)}(\phi; x).$$

We put

$$\zeta_{j,i}^{(i)}(\phi, f_0; s) = \frac{1}{v(\Gamma)} \sum_{x \in \Gamma \backslash V_Q \cap V_i} \frac{\mu(x) f_0(x) v_{j,\mu_l}^{(i)}(\phi; x)}{\prod_{t=1}^n |P_t(x)|^s}$$

and

$$\Phi_j^{(i)}(f_\infty; \pi, \mu_l, s) = \int_{V_i} \prod_{t=1}^n |P_t(y)|^s \Psi_{i,j}^{\pi,\mu_l}(\bar{y}) f_\infty(y) \Omega(y).$$

Now Proposition 1.7 can be formulated as follows:

### PROPOSITION 5.2

When  $\operatorname{Re}(s_1), \dots, \operatorname{Re}(s_n)$  are sufficiently large, we have

$$Z_\phi(s)(f_\infty \otimes f_0) = \sum_{l=1}^d \sum_{j=1}^{r_l} \sum_{i=1}^d \Phi_j^{(i)}(f_\infty; \pi, \mu_l, s) \zeta_{j,i}^{(i)}(\phi, f_0; s).$$

Note that  $\zeta_{j,i}^{(i)}(\phi, f_0; s)$  are Dirichlet series with values in  $W_\pi^M$  and  $\Phi_j^{(i)}(f_\infty; \pi, \mu_l, s)$  are local zeta functions with values in  $\operatorname{End}(W_\pi)$ . Hence the products of  $\Phi_j^{(i)}(f_\infty; \pi, \mu_l, s)$  and  $\zeta_{j,i}^{(i)}(\phi, f_0; s)$  in Proposition 5.2 are well-defined.

**5.3** Let  $(G, \rho, V) = (G, \rho_1 \oplus \rho_2, E \oplus F)$  and assume that  $F$  is a regular subspace. Denote by  $(G, \rho^*, V^*)$  the prehomogeneous vector space dual to  $(G, \rho, V)$  with respect to  $F$ .

In the following we indicate with the superscript  $*$  the notation for  $(G, \rho^*, V^*)$ . For example, we denote by  $P_1^*, \dots, P_n^*$  the basic relative invariants of  $(G, \rho^*, V^*)$ .

Take a relative invariant  $P$  of  $(G, \rho, V)$  with coefficients in  $\mathbb{Q}$  such that  $\phi_P$  defined in the proof of Lemma 4.1 gives a biregular mapping of  $V - S$  onto  $V^* - S^*$ . Since  $\phi_P$  is defined over  $\mathbb{Q}$  and  $G$ -equivariant, we have a one-to-one correspondence of  $P^+$ -orbits in  $V_{\mathbb{R}} - S_{P, \mathbb{R}}$  and those in  $V_{\mathbb{R}}^* - S_{P, \mathbb{R}}^*$ . Hence we have

$$V_{\mathbb{R}}^* - S_{P, \mathbb{R}}^* = \bigcup_{i=1}^v \bigcup_{j=1}^{r_i} V_{ij}^*, \quad V_{ij}^* = \phi_P(V_{ij}) = \rho^*(P^+) x_{ij}^*,$$

$$x_{ij}^* = \phi_P(x_{ij}) = \rho^*(w_j^{(i)}) \phi_P(x_i).$$

Since  $L_{(x_i)}^+ = L_{(x_i^*)}^+$  for  $x_i^* = \phi_P(x_i)$ , we may identify  $X_i = L_0^+/L_{(x_i)}^+$  with  $X_i^* = L_0^+/L_{(x_i^*)}^+$  and the assumption (A-7) holds also for  $(G, \rho^*, V^*)$ . Moreover we have the commutative diagram

$$\begin{array}{ccc} V_i & \xrightarrow{\phi_P} & V_i^* \\ & \searrow & \swarrow \\ & X_i & \end{array}$$

For  $x^* = \phi_P(x) (x \in V_i)$ , it is easy to check the following identity:

$$|\bar{p}^*(\bar{x}^*)|_{ij}^{\mu+\rho} = |\bar{p}(\bar{x})|_{ij}^{\mu+\rho}.$$

If  $x$  is in  $V_i \cap V_Q$  (and hence  $x^*$  is in  $V_i \cap V_Q^*$ ), then we have

$$\mathcal{M}_{x^*}^{(i)} \phi(\bar{y}) = \mathcal{M}_x^{(i)} \phi(\bar{y}) \quad (\bar{y} \in X_i),$$

$$v_{j, \mu_i}^{(i)}(\phi; x^*) = v_{j, \mu_i}^{(i)}(\phi; x).$$

The zeta functions and the local zeta functions associated with  $(G, \rho^*, V^*)$  are defined as follows:

$$\zeta_{j,i}^{*(i)}(\phi, f_0^*; s) = \frac{1}{v(\Gamma)} \sum_{x^* \in \Gamma \backslash V_Q^* \cap V_i^*} \frac{\mu(x^*) f_0^*(x^*) v_{j, \mu_i}^{(i)}(\phi; x^*)}{\prod_{t=1}^n |P_t^*(x^*)|^{s_t}},$$

$$\Phi_j^{*(i)}(f_\infty^*; \pi, \mu_i, s) = \int_{V_i^*} \prod_{t=1}^n |P_t^*(y^*)|^{s_t} \Psi_{i,j}^{\pi, \mu_i}(\bar{y}^*) f_\infty^*(y^*) \Omega^*(y^*).$$

**Theorem 5.3.** For any  $f_\infty \in \mathcal{S}(V_{\mathbb{R}})$  and  $f_\infty^* \in \mathcal{S}(V_{\mathbb{R}}^*)$ , the integrals  $\Phi_j^{(i)}(f_\infty; \pi, \mu, s)$ ,  $\Phi_j^{*(i)}(f_\infty^*; \pi, \mu, s)$   $((\mu, s) \in \mathfrak{a}_{\mathbb{C}}^* \times \mathbb{C}^n)$  converge absolutely, when  $\text{Re}(\langle \mu + \rho, \alpha \rangle) < 0$  for all  $\alpha \in \Delta$  and  $\text{Re}(s_1), \dots, \text{Re}(s_n)$  are sufficiently large. Moreover they have analytic continuations to meromorphic functions of  $(\mu, s)$  in  $\mathfrak{a}_{\mathbb{C}}^* \times \mathbb{C}^n$  and satisfy the functional equation

$$\Phi_j^{(i)}(f_\infty; \pi, \mu, s) = \sum_{i^*=1}^v \sum_{j^*=1}^{r_{i^*}} \Gamma_{j,j^*}^{(i,i^*)}(\mu, s) \Phi_{j^*}^{*(i^*)}(\hat{f}_\infty; \pi, \mu, (s - \lambda)U),$$

where  $\Gamma_{j,j^*}^{(i,i^*)}(\mu, s)$  are meromorphic functions independent of  $f_\infty$  and  $\pi$  having an elementary expression in terms of the gamma function and the exponential function.

*Remark.* In the case where  $(G, \rho, V)$  is of the form  $(G, \rho, V) = R_{\mathbb{C}/\mathbb{R}}(G_1, \rho_1, V_1)$ , where  $(G_1, \rho_1, V_1)$  is an irreducible regular prehomogeneous vector space of commutative parabolic type, Bopp and Rebenthaler ([BR, Théorème 3.2]) obtained a functional equation that is essentially equivalent to the local functional equation in the theorem above.

*Proof.* From (5.4), we have

$$\Phi_j^{(i)}(f_\infty; \pi, \mu, s) = \int_{V_i} \prod_{t=1}^n |P_t(y)|^{s_t} \left\{ \int_K |\bar{p}(k^{-1} \bar{y})|_{ij}^{\mu + \rho} \pi(k) dk \right\} f_\infty(y) \Omega(y).$$

Since  $P_i$ 's are  $K$ -invariant, we obtain

$$\Phi_j^{(i)}(f_\infty; \pi, \mu, s) = \int_{V_{ij}} \prod_{t=1}^n |P_t(y)|^{s_t} |p(y)|^{\mu + \rho} \left\{ \int_K f_\infty(\rho(k)y) \pi(k) dk \right\} \Omega(y). \quad (5.5)$$

Similarly we obtain

$$\begin{aligned} & \Phi_j^{*(i)}(f_\infty^*; \pi, \mu, s) \\ &= \int_{V_{ij}^*} \prod_{t=1}^n |P_t^*(y^*)|^{s_t} |p^*(y^*)|^{\mu + \rho} \left\{ \int_K f_\infty^*(\rho^*(k)y^*) \pi(k) dk \right\} \Omega^*(y^*). \end{aligned} \quad (5.6)$$

Putting  $\mu = \sum_{j=1}^l \mu_j \log(\chi_{n+j} \circ \exp)$  and  $\rho = \sum_{j=1}^l \rho_j \log(\chi_{n+j} \circ \exp)$ , we have

$$\prod_{i=1}^n |P_i(y)|^{s_i - \delta_i} |p(y)|^{\mu + \rho} = \prod_{i=1}^n |P_i(y)|^{s_i - \delta_i - \sum_{j=1}^l (m_{ij} \mu_j + \rho_j)} \prod_{j=1}^l |P_{n+j}(y)|^{\mu_j + \rho_j}.$$

If  $\Lambda \in X(T'_0)_{\mathbb{R}}$  is a dominant weight, then, for some positive integer  $m$ ,  $-m\Lambda$  corresponds to a relative invariant that is regular on  $V - S$  (cf. [V, § 3]). This implies that, if  $\text{Re}(\langle \mu + \rho, \alpha \rangle) < 0$  for all  $\alpha \in \Delta$ , then  $\text{Re}(\mu_j + \rho_j) > 0$  ( $j = 1, \dots, l$ ). Hence the integral (5.5) is absolutely convergent, when  $\text{Re}(\langle \mu + \rho, \alpha \rangle) < 0$  for all  $\alpha \in \Delta$  and  $\text{Re}(s_1), \dots, \text{Re}(s_n)$  are sufficiently large. The proof of the convergence of the integral (5.6) is quite similar. Since any matrix coefficient of

$$\int_K f_\infty(\rho(k)y) \pi(k) dk \quad \left( \text{resp.} \int_K f_\infty^*(\rho^*(k)y) \pi(k) dk \right)$$

is a rapidly decreasing function on  $V_{\mathbb{R}}$  (resp.  $V_{\mathbb{R}}^*$ ), the integrals  $\Phi_j^{(i)}$  (resp.  $\Phi_j^{*(i)}$ ) can be viewed as local zeta functions associated with  $(P, \rho, V)$  of the type considered in [S1, § 5] and have analytic continuations to meromorphic functions on  $a_{\mathbb{C}}^* \times \mathbb{C}^n$  (cf. [BG], [KK], [Sab]). We note further that, for  $u, v \in W_{\pi}$ ,

$$\left\langle \int_K \hat{f}_\infty(\rho(k)y) \pi(k) dk \cdot u, v \right\rangle = \left( \left\langle \int_K f_\infty(\rho(k)y) \pi(k) dk \cdot u, v \right\rangle \right)^{\wedge}.$$

Applying [S1, Theorem 1] to  $(P, \rho, V)$ , we can see that there exist meromorphic



functions  $\Gamma_{j,j^*}^{(i,i^*)}(\mu, s)$  on  $\mathfrak{a}_{\mathbb{C}}^* \times \mathbb{C}^n$  such that the functional equation

$$\langle \Phi_j^{(i)}(f_{\infty}; \pi, \mu, s)u, v \rangle = \sum_{i^*=1}^v \sum_{j^*=1}^{r_i} \Gamma_{j,j^*}^{(i,i^*)}(\mu, s) \langle \Phi_{j^*}^{*(i^*)}(\hat{f}_{\infty}; \pi, \mu, (s-\lambda)U)u, v \rangle$$

holds for all  $u, v \in W_{\pi}$ . This proves the theorem. ■

For  $(\mu, s) \in \mathfrak{a}_{\mathbb{C}}^* \times \mathbb{C}^n$ , put

$$P_{\mu,s}(y) = \prod_{i=1}^n P_i(y)^{s_i - \delta_i - \sum_{j=1}^{r_i} (m_{ij}\mu_j + \rho_j)} \prod_{j=1}^l P_{n+j}(y)^{\mu_j + \rho_j},$$

where  $\mu_j$  and  $\rho_j$  are the same as in the proof of Theorem 5.3 above. Let  $P_F^*$  be the relative invariant of  $(G, \rho^*, V^*)$  introduced just before Lemma 3.2 and  $\chi_F^*$  the rational character of  $G$  corresponding to  $P_F^*$ . Then, by [S1, §3], there exists a polynomial  $b_F(s, \mu)$ , the  $b$ -function of  $(P, \rho, V)$  with respect to  $F$ , satisfying

$$P_F^* \left( y_1, \frac{\partial}{\partial y_2} \right) P_{\mu,s}(y) = b_F(s, \mu) P_{\mu, s+\alpha}(y), \quad (5.7)$$

where  $\alpha = (\alpha_1, \dots, \alpha_n)$  is defined by  $\chi_F^* = \chi_1^{\alpha_1} \cdots \chi_n^{\alpha_n}$ . We can similarly define the  $b$ -function  $b_F^*(s, \mu)$  of  $(P, \rho^*, V^*)$  with respect to  $F^*$ .

Now we are in a position to prove the functional equation of the zeta functions  $\zeta_{j,l}^{(i)}(\phi, f_0; s)$  and  $\zeta_{j,l}^{*(i)}(\phi, f_0^*; s)$ .

**Theorem 5.4.** Assume that  $2\langle \mu_l, \alpha \rangle / \langle \alpha, \alpha \rangle \notin \mathbb{Z}$  for all  $\alpha \in \Sigma$  and  $1 \leq l \leq d$ . Then

(1) the zeta functions  $\zeta_{j,l}^{(i)}(\phi, f_0; s)$  and  $\zeta_{j,l}^{*(i)}(\phi, f_0^*; s)$  can be extended to meromorphic functions of  $s$  in  $D$  and  $D^*$ , respectively (for the definition of  $D$  and  $D^*$ , see §3).

(2) The functions  $b_F(s, \mu_l) \zeta_{j,l}^{(i)}(\phi, f_0; s)$  and  $b_F^*(s, \mu_l) \zeta_{j,l}^{*(i)}(\phi, f_0^*; s)$  are holomorphic functions of  $s$  in  $D$  and  $D^*$ , respectively.

(3) The following functional equation holds for any  $f_0 \in \mathcal{S}(V_Q)$ :

$$\zeta_{j,l}^{*(i^*)}(\phi, \hat{f}_0; (s-\lambda)U) = \sum_{i=1}^v \sum_{j=1}^{r_i} \Gamma_{j,j^*}^{(i,i^*)}(\mu_1, s) \zeta_{j,l}^{(i)}(\phi, f_0; s).$$

*Proof.* (1) and (2): Let the notation be as in §3. For an  $f'_{\infty} \in C_0^{\infty}(V_i)$ , put  $f_{\infty} = P_F^*(x_1, \partial/\partial x_2) f'_{\infty}(x_1, x_2)$ . Then, by Lemma 3.2, we can apply Proposition 3.1 to  $f_{\infty}$  and we see that the function

$$Z_{\phi}(s)(f_{\infty} \otimes f_0) = \sum_{j=1}^{r_i} \sum_{l=1}^d \Phi_j^{(i)}(f_{\infty}; \pi, \mu_l, s) \zeta_{j,l}^{(i)}(\phi, f_0; s)$$

is a holomorphic function of  $s$  in  $D$ . On the other hand

$$\begin{aligned} & \Phi_j^{(i)}(f_{\infty}; \pi, \mu_l, s) \\ &= \int_{V_{i,j}} \prod_{i=1}^n |P_i(y)|^{s_i} |p(y)|^{\mu_l + \rho} \left\{ \int_K P_F^* \left( y_1, \frac{\partial}{\partial y_2} \right) f'_{\infty}(\rho(k)y) \pi(k) dk \right\} \Omega(y), \end{aligned}$$

if the integral in the right hand side of the identity is absolutely convergent. Since

$P_F^*$  is  $K$ -invariant, we have

$$P_F^* \left( y_1, \frac{\partial}{\partial y_2} \right) f'_\infty(\rho(k)y) = P_F^* \left( y_1, \frac{\partial}{\partial y_2} \right) (kf'_\infty)(y), \quad kf'_\infty(y) = f'_\infty(\rho(k)y).$$

Hence, integrating by parts, we obtain

$$\Phi_j^{(i)}(f_\infty; \pi, \mu_l, s) = \pm b_F(s, \mu_l) \Phi_j^{(i)}(f'_\infty; \pi, \mu_l, s + \alpha),$$

where we use the identity (5.7). This identity holds for any  $s$  and  $\mu_l$  by analytic continuation. Thus we see that

$$Z_\phi(s)(f_\infty \otimes f_0) = \sum_{j=1}^{r_l} \sum_{l=1}^d \pm b_F(s, \mu_l) \Phi_j^{(i)}(f'_\infty; \pi, \mu_l, s + \alpha) \zeta_{j,l}^{(i)}(\phi, f_0; s) \quad (5.8)$$

is a holomorphic function in  $D$ .

Now we need the following lemma, whose proof is not hard and is omitted.

**Lemma 5.5.** *Let  $V$  and  $W$  be finite-dimensional  $\mathbb{C}$ -vector spaces. Let  $\Psi: X \rightarrow \text{Hom}(W, V)$  be a  $\text{Hom}(W, V)$ -valued continuous function on a domain  $X$  in  $\mathbb{R}^N$ . Take a basis  $\{w_1, \dots, w_n\}$  ( $n = \dim W$ ) of  $W$  and let  $\Psi_i: X \rightarrow V$  be the function defined by  $\Psi_i(x) = \Psi(x)(w_i)$ . Assume that the functions  $\Psi_1, \dots, \Psi_n$  are linearly independent over  $\mathbb{C}$ . Then there exist  $f_1, \dots, f_n \in C_0^\infty(X)$  such that the linear mapping of  $W$  into  $V^{\oplus n}$  defined by*

$$W \ni w \mapsto \left( \int_X \Psi(x)(w) f_1(x) dx, \dots, \int_X \Psi(x)(w) f_n(x) dx \right) \in V^{\oplus n}$$

is injective.

When  $2\langle \mu_l, \alpha \rangle / \langle \alpha, \alpha \rangle \notin \mathbb{Z}$  ( $1 \leq l \leq d$ ), the lemma can be applied to the function

$$\Psi: V_l \rightarrow \text{Hom}((W_\pi^M)^{\oplus dr_l}, W_\pi)$$

defined by

$$\Psi(x)(v) = \prod_{k=1}^n |P_k(x)|^{s_k + a_k - \delta_k} \sum_{j=1}^{r_l} \sum_{l=1}^d \Psi_{l,j}^{\pi, \mu_l}(\bar{x}) \cdot v_{jl}$$

$$(v = (v_{jl})_{\substack{j=1, \dots, r_l \\ l=1, \dots, d}} \in (W_\pi^M)^{\oplus dr_l}).$$

Hence, by (5.8), we see that the functions  $b_F(s, \mu_l) \zeta_{j,l}^{(i)}(\phi, f_0; s)$  are holomorphic in  $D$ . The holomorphy of  $b_F^*(s, \mu_l) \zeta_{j,l}^{*(i)}(\phi, f_0^*; s)$  can be shown quite similarly.

(3): Now we take  $f'_\infty \in C_0^\infty(V_\pi^*)$  and put  $f_\infty(x_1, x_2) = P_F(x_1, x_2) \cdot \hat{f}'_\infty(x_1, x_2)$ . Then we can apply Proposition 3.1 to  $f_\infty$  and get the functional equation

$$Z_\phi^*((s - \lambda)U)(\hat{f}_\infty \otimes \hat{f}_0) = Z_\phi(s)(f_\infty \otimes f_0) \quad (s \in D).$$

By Proposition 5.2 and Theorem 5.3, we have

$$\sum_{j^*=1}^{r_{l^*}} \sum_{l^*=1}^d \Phi_{j^*}^{*(i^*)}(\hat{f}_\infty; \pi, \mu_{l^*}, (s - \lambda)U) \zeta_{j^*, l^*}^{*(i^*)}(\phi, \hat{f}_0; (s - \lambda)U)$$

$$\begin{aligned}
&= \sum_{i=1}^v \sum_{j=1}^{r_i} \sum_{l=1}^d \Phi_j^{(i)}(f_\infty; \pi, \mu_{l^*}, s) \zeta_{jl}^{(i)}(\phi, f_0; s) \\
&= \sum_{i=1}^v \sum_{j=1}^{r_i} \sum_{l=1}^d \Gamma_{j,j^*}^{(i,i^*)}(\mu_{l^*}, s) \Phi_{j^*}^{*(i^*)}(\hat{f}_\infty; \pi, \mu_{l^*}, (s-\lambda)U) \zeta_{jl}^{(i)}(\phi, f_0; s).
\end{aligned}$$

Therefore

$$\begin{aligned}
&\sum_{j^*=1}^{r_{i^*}} \sum_{l=1}^d \Phi_{j^*}^{*(i^*)}(\hat{f}_\infty; \pi, \mu_{l^*}, (s-\lambda)U) \\
&\quad \times \left( \zeta_{j_l^*}^{*(i^*)}(\phi, \hat{f}_0; (s-\lambda)U) - \sum_{i=1}^v \sum_{j=1}^{r_i} \Gamma_{j,j^*}^{(i,i^*)}(\mu_{l^*}, s) \zeta_{jl}^{(i)}(\phi, f_0; s) \right) = 0.
\end{aligned}$$

By an argument based upon Lemma 5.5, we see that the functional equation

$$\zeta_{j_l^*}^{*(i^*)}(\phi, \hat{f}_0; (s-\lambda)U) = \sum_{i=1}^v \sum_{j=1}^{r_i} \Gamma_{j,j^*}^{(i,i^*)}(\mu_{l^*}, s) \zeta_{jl}^{(i)}(\phi, f_0; s).$$

holds for any  $s \in D$ . ■

*Remarks.* (1) As we mentioned in the introduction, the functional equation of zeta functions are based on local functional equations, the existence of  $b$ -functions and the functional equations of the zeta integrals. In the case considered above, the local functional equation (Theorem 5.3) and the  $b$ -function (5.7) are reduced to the usual local functional equations and the  $b$ -functions of the prehomogeneous vector space  $(P, \rho, V)$ .

(2) By the results of Oshima [O], Proposition 5.1 can be extended to the case where the symmetric spaces  $X_i = L_0^+ / L_{(x_0)}^+$  are not necessarily of  $K_e$ -type. By a similar argument based on the extended version of Proposition 5.1, we can extend the functional equations of zeta functions attached to automorphic forms to more general prehomogeneous vector spaces with symmetric structure not of  $K_e$ -type. In the general case,  $P$  is not necessarily minimal parabolic, and the functional equations are reduced to the local functional equations of local zeta functions attached to representations of the Levi part of  $P$ , which belong to the class of local zeta functions studied in [S6, § 3].

## 6. Examples of functional equations

In this section, we present some examples of zeta functions, to which we can apply the results in the previous section. We retain the notation in § 5.

### 6.1 Zeta functions of Maass attached to positive definite quadratic forms

First we consider (a special case of) the zeta functions studied by Maass in [M1], [M2] and [M4]. Let  $SO(m)$  be the special orthogonal group of the quadratic form  $v_1^2 + \cdots + v_m^2$ . The prehomogeneous vector space to be considered is  $(G, \rho, V) = (SO(m) \times GL(n), \rho, M(m, n))$  ( $m \geq n, m \geq 3$ ), where the representation  $\rho$  is given by  $\rho(k, g)x = kxg^{-1}$  ( $k \in SO(m), g \in GL(n), x \in M(m, n)$ ). The singular set  $S$  is given by

$$S = \{x \in V \mid \det(x^t x) = 0\}$$

and  $P(x) = \det(x'x)$  is the basic relative invariant. We can naturally regard  $(G, \rho, V)$  as a prehomogeneous vector space defined over  $\mathbb{Q}$ . Then  $V_{\mathbb{R}} - S_{\mathbb{R}} = \{x \in M(m, n; \mathbb{R}) \mid \text{rank } x = n\}$  and this is a single  $G^+$ -orbit. Put  $x_0 = \begin{pmatrix} I_n \\ 0^{(m-n, n)} \end{pmatrix}$ .

Consider the symmetric structure defined by putting  $L = GL(n)$  and  $U = SO(m)$  (cf. §4.1 Example 3). Then  $L_0^+ = SL(n) := \mathbf{SL}(n)_{\mathbb{R}}$ ,  $L_{(x_0)}^+ = SO(n) := \mathbf{SO}(n)_{\mathbb{R}}$  and  $v = 1$ . The corresponding symmetric space is  $X_1 = SL(n)/SO(n)$ . We identify  $X_1$  with the space of positive definite symmetric matrices of size  $n$  with determinant 1.

Let  $\mathfrak{a}$  be the set of real diagonal matrices with trace 0. The complexification  $\mathfrak{a}_{\mathbb{C}}$  is a Cartan subalgebra of  $\mathfrak{g}_{0\mathbb{C}}$ . We define  $\Lambda_i \in \mathfrak{a}_{\mathbb{C}}^*$  ( $1 \leq i \leq n-1$ ) by

$$\Lambda_i \left( \begin{pmatrix} a_1 & & 0 \\ & \ddots & \\ 0 & & a_n \end{pmatrix} \right) = a_1 + \cdots + a_i.$$

Then  $\{\Lambda_1, \dots, \Lambda_{n-1}\}$  forms a basis of  $\mathfrak{a}_{\mathbb{C}}^*$  (the fundamental dominant weights). We put  $\lambda = \sum_{i=1}^{n-1} \lambda_i \Lambda_i \in \mathfrak{a}_{\mathbb{C}}^*$  ( $\lambda_i \in \mathbb{C}$ ). We denote by  $\chi = \chi_{\lambda}$  the infinitesimal character of  $\mathcal{X}(X_1) = \mathcal{D}(X_1)$  defined by (5.2). The character  $\chi$  canonically induces a character of  $\mathcal{X}(L_0^+)$ , which we denote by the same symbol  $\chi$ . Let  $\pi$  be an irreducible unitary representation of  $K = SO(n)$  and  $W_{\pi}$  its representation space. Put

$$\Psi_{\lambda}(\bar{y}) = \int_{SO(n)} \prod_{i=1}^{n-1} d_i({}^t k \bar{y} k)^{-(\lambda_i + 1)/2} \pi(k) dk, \quad \bar{y} = \frac{{}^t y y}{(\det {}^t y y)^{1/n}}, \quad (6.1)$$

where  $d_i(A)$  is the determinant of the upper left  $i$  by  $i$  block of a square matrix  $A$ . In the present case, Proposition 5.1 is the integral representation of  $K$ -finite eigenfunctions on  $X$  due to Helgason ([Hel2, Corollary 7.4]) and any element in  $\mathcal{E}(X_1; \pi, \chi)$  is of the form  $\Psi_{\lambda}(\bar{y})v$  ( $v \in W_{\pi}$ ).

Put  $\Gamma = \mathbf{SL}(n)_{\mathbb{Z}}$ . Since an automorphic form  $\phi \in \mathcal{A}(L_0^+/\Gamma; \pi, \chi)$  is slowly increasing and  $X_1$  is a riemannian symmetric space, the condition (A-3) is satisfied. By (6.1), for each  $x \in V_{\mathbb{Q}} - S_{\mathbb{Q}}$ , one can find a unique  $v(\phi, x) \in W_{\pi}^M$  such that

$$\mathcal{M}_x \phi(y) := \int_{SO(n)} \phi(h, k h_x^{-1}) dk = \Psi_{\lambda}(\bar{y}) v(\phi, x).$$

In particular, if  $\pi$  is the trivial representation  $\pi_0$  and  $\phi$  is  $SO(n)$ -invariant, then  $\Psi_{\lambda}(\bar{y}) = \omega_{\lambda}(\bar{y})$ , the zonal spherical function of  $SL(n)$ . Any automorphic form  $\phi \in \mathcal{A}(L_0^+/\Gamma; \pi_0, \chi)$  can be viewed as a  $\Gamma$ -invariant function on  $X_1$  and we have  $v(\phi, x) = \phi(\bar{x}^{-1})$ .

Now return to the case of general  $\pi$ . For a  $\Gamma$ -invariant  $f_0 \in \mathcal{S}(V_{\mathbb{Q}})$  and an  $f_{\infty} \in \mathcal{S}(V_{\mathbb{R}})$ , the zeta function and the local zeta function are given as follows:

$$\zeta(\phi, f_0; s) = \sum_{\substack{x \in V_0/\Gamma \\ \text{rank } x = n}} \frac{f_0(x) v(\phi; x)}{(\det {}^t x x)^s},$$

$$\Phi(f_{\infty}; \pi, \lambda, s) = \int_{V_{\mathbb{R}} - S_{\mathbb{R}}} (\det {}^t x x)^{s-m/2} \Psi_{\lambda}(\bar{x}) f_{\infty}(x) dx.$$

By the proof of Theorem 5.3 (or by the integral representation (6.1)), the local functional equation satisfied by  $\Phi(f_\infty; \pi, \lambda, s)$  can be reduced to the functional equation of

$$\xi^{(\infty)}(f_\infty; s, \lambda) = \int_{\mathbf{V}_R - \mathbf{S}_R} \prod_{i=1}^{n-1} d_i(\bar{x})^{-(\lambda_i+1)/2} (\det' x x)^{s-m/2} f_\infty(x) dx,$$

which is the local zeta function of the prehomogeneous vector space  $(\mathbf{SO}(m) \times \mathbf{B}(n), \rho, \mathbf{M}(m, n))$ , where  $\mathbf{B}(n)$  is the subgroup of upper triangular matrices in  $\mathbf{GL}(n)$ .

Using the functional equation of  $\xi^{(\infty)}(f_\infty; s, \lambda)$  calculated in [S, pp. 155–156], we obtain the following explicit functional equation, which connects  $\zeta(\phi, f_0; s)$  to  $\zeta(\check{\phi}, \hat{f}_0; s)$ , where  $\check{\phi}(h) = \phi({}^t h^{-1})$ . Note that  $\check{\phi} \in \mathcal{A}(L_0^+/\Gamma; \pi, \chi_\chi)$  where  $\check{\lambda} = \sum_{i=1}^{n-1} \lambda_{n-i} \Lambda_i$ .

**Theorem 6.1.** (1) *The zeta function  $\zeta(\phi, f_0; s)$  has an analytic continuation to a meromorphic function of  $s$  in  $\mathbb{C}$ .*

(2) *Set*

$$\Lambda(\phi, f_0; s) = \pi^{-ns} \prod_{i=1}^n \Gamma\left(s + \frac{1}{2n} \sum_{j=1}^{n-1} j \lambda_j - \frac{1}{2} \sum_{j=i}^{n-1} \lambda_j - \frac{n-1}{4}\right) \cdot \zeta(\phi, f_0; s),$$

$$\Lambda(\check{\phi}, \hat{f}_0; s) = \pi^{-ns} \prod_{i=1}^n \Gamma\left(s + \frac{1}{2n} \sum_{j=1}^{n-1} j \lambda_{n-j} - \frac{1}{2} \sum_{j=i}^{n-1} \lambda_{n-j} - \frac{n-1}{4}\right) \cdot \zeta(\check{\phi}, \hat{f}_0; s).$$

Then

$$\Lambda(\check{\phi}, \hat{f}_0; s) = \Lambda\left(\phi, f_0; \frac{m}{2} - s\right).$$

**Remarks.** (1) In [M1], [M2], and [M4], Maass considered the case where  $\pi$  is the trivial representation  $\pi_0$  of  $\mathbf{SO}(m)$ .

(2) The zeta function  $\zeta(\phi, f_0, s)$  is closely related to the Eisenstein series of  $\mathbf{SL}(n)$  corresponding to the standard parabolic subgroup defined by the partition  $(n, m-n)$  (cf. [T, Chapter IV, § 4.5]).

(3) If  $n, m-n \neq 2$ , then the results in this section can be generalized to arbitrary  $\mathbb{Q}$ -forms of  $(\mathbf{SO}(m) \times \mathbf{GL}(n), \rho, \mathbf{V})$  including the case where the quadratic form defining  $\mathbf{SO}(m)$  is indefinite.

## 6.2 Zeta functions with Maass cusp forms

Let  $\mathbf{G} = \mathbf{GL}(2)$ ,  $\mathbf{V} = \mathbf{Sym}(2) = \{x \in \mathbf{M}(2) \mid {}^t x = x\}$ , and define a rational representation  $\rho$  of  $\mathbf{G}$  on  $\mathbf{V}$  by  $\rho(g)x = gx'g$ . Then,  $(\mathbf{G}, \rho, \mathbf{V})$  is a prehomogeneous vector space with singular set  $\mathbf{S} = \{x \in \mathbf{V} \mid \det x = 0\}$ . The prehomogeneous vector space admits a symmetric structure defined by  $\mathbf{L} = \mathbf{G}$  and  $\mathbf{U} = \{1\}$  (cf. § 4.1, Example 1).

Let  $\mathcal{H}$  be the upper half plane and  $\Phi: \mathcal{H} \rightarrow \mathbb{C}$  be a Maass cusp form with respect to  $\Gamma = \mathbf{SL}(2)_{\mathbb{Z}}$ . Then  $\Phi$  has the following properties:

$$\Phi(\gamma \cdot z) = \Phi(z) \quad (\forall \gamma \in \Gamma); \quad (6.2)$$

$$\Delta \Phi(z) = \lambda(\lambda-1)\Phi(z), \quad \Delta = y^2 \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right); \quad (6.3)$$

for any  $r > 0$ ,  $\Phi(z) = O(e^{-r})$  as  $y \rightarrow \infty$ . (6.4)

Here we note that the estimate in (6.4) is uniform with respect to  $x = \operatorname{Re}(z)$ .

For  $g \in L_0^+ = SL(2) = \mathbf{SL}(2)_{\mathbb{R}}$ , we put  $\phi(g) = \Phi(g^{-1} \cdot \sqrt{-1})$ . Let

$$g = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} y^{-1/2} & 0 \\ 0 & y^{1/2} \end{pmatrix} \begin{pmatrix} 1 & -x \\ 0 & 1 \end{pmatrix}$$

be the Iwasawa decomposition of  $g \in L_0^+$ . Then we have  $\phi(g) = \Phi(x + y\sqrt{-1})$ . For simplicity, we write

$$k_\theta = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}, \quad a_y = \begin{pmatrix} y^{-1/2} & 0 \\ 0 & y^{1/2} \end{pmatrix}, \quad n_x = \begin{pmatrix} 1 & -x \\ 0 & 1 \end{pmatrix}.$$

Then  $\phi(g)$  satisfies the following conditions:

$$\phi(k_\theta g \gamma) = \phi(g) \quad (\theta \in \mathbb{R}, \gamma \in \Gamma),$$

$$\mathcal{C}\phi = \lambda(1 - \lambda)\phi,$$

where  $\mathcal{C}$  is the Casimir operator of  $L_0^+$ .

We consider the natural  $\mathbb{Q}$ -structure on  $(G, \rho, V)$ . Take a lattice  $L$  in  $V_{\mathbb{Q}} = \mathbf{Sym}(2, \mathbb{Q})$  that is stable under  $\Gamma$ . Let  $f_L$  be the characteristic function of  $L$ . For  $f_\infty \in \mathcal{S}(V_{\mathbb{R}})$ , set

$$\Theta(\phi, L, f_\infty) = \int_{L_0^+/\Gamma} \phi(g) \sum_{x \in L-S} f_\infty(gx'g) dg.$$

Then we have

$$Z_\phi(s)(f_\infty \otimes f_L) = \int_0^{+\infty} t^{2s} \Theta(\phi, L, f_\infty) \frac{dt}{t},$$

where  $f_\infty^t(x) = f_\infty(tx)$ . Though the prehomogeneous vector space  $(G, \rho, V)$  does not satisfy the latter half of the condition (A-1), it follows from the estimate (6.4) that the zeta integral  $Z_\phi(s)(f_\infty \otimes f_L)$  is absolutely convergent for  $\operatorname{Re}(s) > \frac{3}{2}$ . Note that the convergence of the zeta integral is sufficient for the application of the results in §5.

Decompose  $V_{\mathbb{R}} - S_{\mathbb{R}}$  as follows:

$$V_{\mathbb{R}} - S_{\mathbb{R}} = V_+ \cup V_-, \quad V_+ = \{x \in V_{\mathbb{R}} | \det x > 0\}, \quad V_- = \{x \in V_{\mathbb{R}} | \det x < 0\}.$$

For  $I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  and  $J_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ , put  $H_+ = SO(I_2)$ ,  $H_- = SO(J_2)$ :

$$H_+ = SO(2) = K = \{k_\theta | 0 \leq \theta < 2\pi\},$$

$$H_- = SO(1, 1) = \left\{ \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \middle| a \in \mathbb{R}^\times \right\}.$$

We normalize the Haar measures  $da$  on  $L^+$  and  $da_-$  on  $H_-$  by

Then we have the following integration formulas:

$$\begin{aligned}
 \int_0^\infty \frac{dt}{t} \int_{L_0^+} f(t, g) dg &= \int_{V_+^p} \frac{dx}{|\det x|^{3/2}} \int_{H_+} f(t_x, g_x k) d\mu_+(k) \\
 &= \int_{V_+^n} \frac{dx}{|\det x|^{3/2}} \int_{H_+} f(t_x, g_x k) d\mu_+(k) \\
 &= \int_{V_-} \frac{dx}{|\det x|^{3/2}} \int_{H_-} f(t_x, g_x h) d\mu_-(h),
 \end{aligned} \tag{6.5}$$

where  $V_+^p$  (resp.  $V_+^n$ ) is the set of positive (resp. negative) definite symmetric matrices in  $V_+$ . Here  $t_x$  and  $g_x$  are defined as follows.

Let  $X_+ = \{x \in V_+^p | \det x = 1\}$  and  $X_- = \{x \in V_- | \det x = -1\}$ . For  $x \in V_{\mathbb{R}} - S_{\mathbb{R}}$  we take  $t_x \in \mathbb{R}_+^*$  and  $g_x \in L_0^+$  such that

$$x = \begin{cases} t_x(g_x I_2 {}^t g_x) & \text{if } x \in V_+^p, \\ -t_x(g_x I_2 {}^t g_x) & \text{if } x \in V_+^n, \\ t_x(g_x J_2 {}^t g_x) & \text{if } x \in V_-. \end{cases}$$

We put  $\bar{x} = g_x I_2 {}^t g_x$  or  $g_x J_2 {}^t g_x$  according as  $\det x > 0$  or  $\det x < 0$ .

For an  $x \in L_{\pm} = L \cap V_{\pm}$  and a  $y \in V_{\pm}$ , consider the mean value

$$\mathcal{M}_x \phi(\bar{y}) = \int_{H_{\pm}/g_x^{-1} \Gamma_x g_x} \phi(g_y h g_x^{-1}) d\mu_{\pm}(h).$$

Then  $\mathcal{M}_x \phi(\bar{y})$  is absolutely convergent and belongs to the space  $\mathcal{E}(X_{\pm}; \lambda)$  of functions  $\psi$  satisfying

$$\psi(k\bar{x}) = \psi(\bar{x}) \quad (k \in K, \bar{x} \in X_{\pm}),$$

$$\Delta_{\pm} \psi = \lambda(1 - \lambda)\psi,$$

where  $\Delta_{\pm}$  is the  $L_0^+$ -invariant differential operator on  $X_{\pm}$  induced by the Casimir operator  $\mathcal{C}$ .

The space  $\mathcal{E}(X_{\pm}; \lambda)$  can be determined explicitly as follows (cf. [Ro, Theorems 6 and 10]. See also [Se]). In this case, we can remove the restriction on  $\lambda$  in Proposition 5.1.

**Lemma 6.2.** (1) Assume that  $\bar{x}$  is in  $X_+$ . Let  $e^{\alpha}$  and  $e^{-\alpha}$  be the eigenvalues of  $\bar{x}$  ( $\alpha \in \mathbb{R}$ ). Then any spherical function in  $\mathcal{E}(X_+; \lambda)$  is a constant multiple of

$$\Psi_{\lambda}(\bar{x}) = P_{-\lambda}(\cosh(\alpha)),$$

where  $P_{-\lambda}(z)$  is the Legendre function and is defined by the Gauss hypergeometric function as follows:

$$P_{-\lambda}(z) = F\left(\lambda, 1 - \lambda; 1; \frac{1 - z}{2}\right).$$

(2) Assume that  $\bar{x}$  is in  $X_-$ . Let  $e^\alpha$  (resp.  $-e^{-\alpha}$ ) be the positive (resp. negative) eigenvalue of  $\bar{x}$  ( $\alpha \in \mathbb{R}$ ). Then any spherical function in  $\mathcal{E}(X_-; \lambda)$  is a linear combination of

$$\Psi_\lambda^{(0)}(\bar{x}) = \cosh(-\alpha)^{-1/2} P_{-1/2}^{\lambda-(1/2)}(\tanh(-\alpha)),$$

$$\Psi_\lambda^{(1)}(\bar{x}) = \cosh(\alpha)^{-1/2} P_{-1/2}^{\lambda-(1/2)}(\tanh(\alpha)),$$

where  $P_{-1/2}^{\lambda-(1/2)}(z)$  is the associated Legendre function and is defined by the Gauss hypergeometric function as follows:

$$P_{-1/2}^{\lambda-(1/2)}(z) = \frac{1}{\Gamma(\frac{3}{2}-\lambda)} \left( \frac{1+z}{1-z} \right)^{(2\lambda-1)/4} F\left(\frac{1}{2}, \frac{3}{2}-\lambda; \frac{1-z}{2}\right).$$

(3) The functions  $\Psi_\lambda(\bar{x})$ ,  $\Psi_\lambda^{(0)}(\bar{x})$  and  $\Psi_\lambda^{(1)}(\bar{x})$  have the following integral representations, which are special cases of (5.4):

$$\Psi_\lambda(\bar{x}) = \frac{1}{2\pi} \int_0^{2\pi} (k_\theta \cdot \bar{x})_{11}^{-\lambda} d\theta,$$

$$\Psi_\lambda^{(0)}(\bar{x}) = \frac{1}{\sqrt{2\pi} \Gamma(1-\lambda)} \int_0^{2\pi} |(k_\theta \cdot \bar{x})_{11}|_+^{-\lambda} d\theta,$$

$$\Psi_\lambda^{(1)}(\bar{x}) = \frac{1}{\sqrt{2\pi} \Gamma(1-\lambda)} \int_0^{2\pi} |(k_\theta \cdot \bar{x})_{11}|_-^{-\lambda} d\theta,$$

where  $A_{11}$  denotes the (1,1)-entry of the matrix  $A$ ,  $|t|_+ = t$  or 0 according as  $t > 0$  or  $t \leq 0$ , and  $|t|_- = -|t|_+$ .

In the proof of Lemma 6.3 below, we need the following formulas for the special values of the Legendre functions.

$$P_{-\lambda}(1) = 1, \quad P_{-1/2}^{\lambda-(1/2)}(0) = \frac{2^{\lambda-(1/2)} \sqrt{\pi}}{\Gamma\left(1 - \frac{\lambda}{2}\right)}. \quad (6.6)$$

Lemma 6.3. (1) If  $x \in V_+$ , then

$$\mathcal{M}_x \phi(\bar{y}) = \frac{\pi}{\varepsilon(x)} \cdot \Phi(z_x) \Psi_\lambda(\bar{y}),$$

where  $\varepsilon(x) = \#(\Gamma_x)$  and  $z_x = g_x \cdot \sqrt{-1}$ .

(2) If  $x \in V_-$ , then

$$\frac{\mathcal{M}_x \phi(\bar{y}) + \mathcal{M}_{-x} \phi(\bar{y})}{2} = \frac{\Gamma\left(1 - \frac{\lambda}{2}\right)^2}{2^{\lambda+(1/2)} \sqrt{\pi}} \{ \mathcal{M}_x \phi(J_2) (\Psi_\lambda^{(0)}(\bar{y}) + \Psi_\lambda^{(1)}(\bar{y})) \}.$$

Proof. (1) By Lemma 6.2 (1),  $\mathcal{M}_x \phi(\bar{y})$  is a constant multiple of  $\Psi_\lambda(\bar{y})$ . Since  $\Psi_\lambda(I_2) = 1$  (see (6.6)), We have

$$\mathcal{M}_x \phi(\bar{y}) = \mathcal{M}_x \phi(I_2) \Psi_\lambda(\bar{y}).$$



By the definition of  $\mathcal{M}_x \phi(\bar{y})$ , we obtain

$$\mathcal{M}_x \phi(I_2) = \int_{H_+/g_x^{-1}\Gamma_x g_x} \phi(kg_x^{-1}) d\mu_+(k) = \frac{1}{2} \varepsilon(x)^{-1} \phi(g_x^{-1}) \int_0^{2\pi} d\theta = \frac{\pi}{\varepsilon(x)} \Phi(z_x).$$

(2) By (1) and Lemma 6.2 (2),  $\mathcal{M}_x \phi(\bar{y})$  can be written as

$$\mathcal{M}_x \phi(\bar{y}) = c^{(0)}(\phi; x) \Psi_\lambda^{(0)}(\bar{y}) + c^{(1)}(\phi; x) \Psi_\lambda^{(1)}(\bar{y}).$$

Note that  $g_{-x} = g_x w$ , where  $w = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ . Hence

$$\begin{aligned} \mathcal{M}_{-x} \phi(-y) &= \int_{H_-/w^{-1}g_x^{-1}\Gamma_x g_x w} \phi(g_y whw^{-1}g_x^{-1}) d\mu_-(h) \\ &= \int_{H_-/g_x^{-1}\Gamma_x g_x} \phi(g_y hg_x^{-1}) d\mu_-(h) = \mathcal{M}_x \phi(\bar{y}). \end{aligned}$$

This implies that  $c^{(0)}(\phi; x) = c^{(1)}(\phi; -x)$  and

$$\mathcal{M}_x \phi(\bar{y}) + \mathcal{M}_{-x} \phi(\bar{y}) = (c^{(0)}(\phi; x) + c^{(1)}(\phi; x))(\Psi_\lambda^{(0)}(\bar{y}) + \Psi_\lambda^{(1)}(\bar{y})).$$

Putting  $y = J_2$ , we have by (6.6)

$$\frac{\mathcal{M}_x \phi(J_2) + \mathcal{M}_{-x} \phi(J_2)}{2} = \frac{2^{\lambda-(1/2)} \sqrt{\pi}}{\Gamma\left(1 - \frac{\lambda}{2}\right)^2} (c^{(0)}(\phi; x) + c^{(1)}(\phi; x)).$$

Since

$$\mathcal{M}_{-x} \phi(J_2) = \mathcal{M}_x \phi(-J_2) = \mathcal{M}_x \phi(w \cdot J_2) = \mathcal{M}_x \phi(J_2),$$

we have

$$c^{(0)}(\phi; x) + c^{(1)}(\phi; x) = \frac{\Gamma\left(1 - \frac{\lambda}{2}\right)^2}{2^{\lambda-(1/2)} \sqrt{\pi}} \mathcal{M}_x \phi(J_2).$$

This proves the lemma. ■

*Remark.* Since  $\phi$  is  $K$ -invariant, the representation of  $SL(2)$  generated by  $\phi$  does not belong to the discrete series. Hence  $\lambda$  cannot be a positive integer and  $\Gamma(1 - \lambda/2) \neq \infty$ .

Now we can define zeta functions  $\zeta_\pm(\phi, L; s)$  and local zeta functions  $\Phi_\pm(f; \lambda, s)$  attached to  $\Phi$  as follows:

$$\zeta_+(\phi, L; s) = \sum_{x \in \Gamma \backslash L_+^*} \frac{\Phi(z_x)}{\varepsilon(x) |\det x|^s},$$

$$\zeta_-(\phi, L; s) = \sum_{x \in \Gamma \backslash L_-} \frac{\mathcal{M}_x \phi(J_2)}{|\det x|^s},$$

$$\Phi_\pm(f_\infty; \lambda, s) = \int_{V_\pm} |\det x|^{s-3/2} \Psi_\lambda^\pm(\bar{x}) f(x) dx.$$

$$\Psi_{\lambda}^{+}(\bar{x}) = \Psi_{\lambda}(\bar{x}) \quad (x \in V_{+}),$$

$$\Psi_{\lambda}^{-}(\bar{x}) = \Psi_{\lambda}^{(0)}(\bar{x}) + \Psi_{\lambda}^{(1)}(\bar{x}) \quad (x \in V_{-}).$$

The following proposition is an explicit version of the integral representation given by Proposition. 1.7.

#### PROPOSITION 6.4

For  $\text{Re}(s) > \frac{3}{2}$ , we have

$$Z_{\phi}(s)(f_{\infty} \otimes f_L) = \frac{1}{\pi} \cdot \zeta_{+}(\phi, L; s) \Phi_{+}(f_{\infty}; \lambda, s) \\ + \frac{\Gamma\left(1 - \frac{\lambda}{2}\right)^2}{2^{\lambda + (1/2)} \sqrt{\pi}} \cdot \zeta_{-}(\phi, L; s) \Phi_{-}(f_{\infty}; \lambda, s).$$

We identify the dual vector space  $V_{\mathbb{R}}^{*}$  with  $V_{\mathbb{R}}$  via the inner product  $\langle x, x^{*} \rangle = \text{tr}(xwx^{*}w^{-1})$ , where  $w = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ . By the cuspidal property of  $\phi$  and the fact that  $\phi(g^{-1}) = \phi(wgw^{-1}) = \phi(g)$ , we have

$$v(L)\Theta(\phi, L, f_{\infty}) = \Theta(\phi, L^{*}, \hat{f}_{\infty}),$$

where  $L^{*}$  is the lattice dual to  $L$  and

$$\hat{f}_{\infty}(x^{*}) = \int_{V_{\mathbb{R}}} f_{\infty}(x) \exp(2\pi\sqrt{-1}\langle x, x^{*} \rangle) dx.$$

Using this theta transformation formula, we can derive the following proposition.

#### PROPOSITION 6.5

The zeta integral  $Z_{\phi}(s)(f_{\infty} \otimes f_L)$  has an analytic continuation to an entire function of  $s$  in  $\mathbb{C}$  and satisfies the functional equation

$$v(L)Z_{\phi}(s)(f_{\infty} \otimes f_L) = Z_{\phi}(\frac{3}{2} - s)(\hat{f}_{\infty} \otimes f_{L^{*}}),$$

$$\text{where } v(L) = \int_{V_{\mathbb{R}}/L} dx.$$

*Remarks.* By the cuspidal property of  $\phi$ , we can prove the functional equation of the zeta integral without the vanishing of  $f_{\infty}$  and on  $\hat{f}_{\infty}$  on the singular sets (compare to Proposition 3.1). This implies that the zeta functions have analytic continuations to entire functions.

**PROPOSITION 6.6**

$$\begin{pmatrix} \Phi_+(\hat{f}_\infty; \lambda, s) \\ \Phi_-(\hat{f}_\infty; \lambda, s) \end{pmatrix} = 2^{1-2s} \pi^{1/2-2s} \Gamma\left(s + \frac{\lambda-1}{2}\right) \Gamma\left(s - \frac{\lambda}{2}\right) \\ \times \begin{pmatrix} \cos(\pi s) & \frac{\Gamma(1-\lambda)}{\sqrt{2\pi}} \cos\left(\frac{\pi\lambda}{2}\right) \\ \frac{\sqrt{2\pi}}{\Gamma(1-\lambda)} \sin\left(\frac{\pi\lambda}{2}\right) & \sin(\pi s) \end{pmatrix} \begin{pmatrix} \Phi_+(f_\infty; \lambda, \frac{3}{2}-s) \\ \Phi_-(f_\infty; \lambda, \frac{3}{2}-s) \end{pmatrix}.$$

As we can see from the proof of Theorem 5.3, the proposition above is reduced to the functional equation of

$$\xi_+^{(\infty)}(f_\infty; \lambda, s) = \int_{V_+} |x_{11}|^{-\lambda} |\det x|^{s+(\lambda-3)/2} f(x) dx, \\ \xi_-^{(\infty)}(f_\infty; \lambda, s) = \int_{V_-} |x_{11}|^{-\lambda} |\det x|^{s+(\lambda-3)/2} f(x) dx,$$

which are the local zeta functions of the prehomogeneous vector space obtained by restricting the representation  $\rho$  to the Borel subgroup of  $\mathbf{GL}(2)$ . In fact, by the integral representations of  $\Psi_\lambda^\pm(\bar{x})$  given in Lemma 6.2, we obtain the following relation between  $\Phi_\pm(f_\infty; \lambda, s)$  and  $\xi_\pm^{(\infty)}(f_{\infty,0}; \lambda, s)$ .

$$\Phi_+(f_\infty; \lambda, s) = \frac{1}{2\pi} \cdot \xi_+^{(\infty)}(f_{\infty,0}; \lambda, s), \\ \Phi_-(f_\infty; \lambda, s) = \frac{1}{\sqrt{2\pi} \Gamma(1-\lambda)} \cdot \xi_-^{(\infty)}(f_{\infty,0}; \lambda, s),$$

where  $f_{\infty,0}(x) = \int_0^{2\pi} f_\infty(k_\theta \cdot x) d\theta$ . The functional equation of  $\xi_\pm^{(\infty)}(f_{\infty,0}; \lambda, s)$  is given in [Sh, Lemma 1 (i)] and Proposition 6.6 is its immediate consequence.

Summing up the results above, we obtain the following explicit form of the functional equation given in Theorem 5.4.

**Theorem 6.7** *The zeta functions  $\zeta_\pm(\phi, L; s)$  have analytic continuations to entire functions of  $s$  and satisfy the functional equation*

$$v(L) \begin{pmatrix} \zeta_+(\phi, L; \frac{3}{2}-s) \\ \zeta_-(\phi, L; \frac{3}{2}-s) \end{pmatrix} = 2^{1-2s} \pi^{1/2-2s} \Gamma\left(s + \frac{\lambda-1}{2}\right) \Gamma\left(s - \frac{\lambda}{2}\right) \\ \times \begin{pmatrix} \cos(\pi s) & \frac{\pi \Gamma\left(1 - \frac{\lambda}{2}\right)^2}{2^\lambda \Gamma(1-\lambda)} \sin\left(\frac{\pi\lambda}{2}\right) \\ \frac{2^\lambda \Gamma(1-\lambda)}{\pi \Gamma\left(1 - \frac{\lambda}{2}\right)^2} \cos\left(\frac{\pi\lambda}{2}\right) & \sin(\pi s) \end{pmatrix} \begin{pmatrix} \zeta_+(\phi, L^*; s) \\ \zeta_-(\phi, L^*; s) \end{pmatrix}.$$

### Remarks on further examples.

(1) The results in §5 apply to Example 1 in §4.1, even if  $n \geq 3$ . In this case, the functional equation can be reduced to the functional equation of the local zeta functions attached to the prehomogeneous vector space  $(\mathbf{B}(n), \rho, \text{Sym}(n))$ , whose explicit formula is found in [S5, Theorem 3.2] (cf. [S3, Theorem 8]). Here we denote by  $\mathbf{B}(n)$  the group of upper triangular matrices of size  $n$ , the Borel subgroup of  $\text{GL}(n)$ . We also note that the zeta functions studied in [S3, §4] can be viewed as the zeta functions attached to  $\phi =$  the Eisenstein series for a minimal parabolic subgroup.

(2) The zeta functions  $\zeta_{\pm}(\phi, L; s)$  admit an interpretation as the Mellin transform of the modular form of weight  $\frac{1}{2}$  obtained from  $\phi$  by the Maass correspondence (cf. [KS], [Hej]). From this point of view, the zeta functions studied in [M3] are a generalization of  $\zeta_{\pm}(\phi, L; s)$  to the zeta functions attached to automorphic forms on orthogonal groups with trivial  $K$ -type. Though the symmetric structure corresponding to the Maass case is in general not of  $K_e$  type, the necessary harmonic analysis was developed by Rossman [Ro] (or Sekiguchi [Se]) and we can generalize Maass' results further to automorphic forms with non-trivial  $K$ -type, as Hejhal did for  $\text{SL}(2)$  in [Hej].

### Acknowledgements

The main part of this paper was written during the author's stay in Strasbourg in the spring of 1990. The author would like to express his sincere gratitude to Département de Mathématiques de Université Louis Pasteur, in particular to Professors H Rubenthaler and G Schiffmann, for their hospitality. Thanks are also due to Egami and Arakawa. Discussions with them in 1983 on the works of Maass ([M3]) and Hejhal ([Hej]) were the starting point of the present investigation.

### References

- [B1] van den Ban E P, Invariant differential operators on a semisimple symmetric space and finite multiplicities in a Plancherel formula, *Ark. Mat.* **25** (1987) 175–187
- [B2] van den Ban E P, Asymptotic behaviour of matrix coefficients related to reductive symmetric spaces, *Proc. K. Ned. Akad. Wet.* **A90** (1987) 225–249
- [BG] Bernstein I N and Gelfand S I, Meromorphic properties of the functions  $P^{\lambda}$ , *Funct. Anal. Appl.* **3** (1969) 68–69
- [BR] Bopp N and Rubenthaler H, Fonction zêta associée à la série principe sphérique de certaines espaces symétriques, *C. R. Acad. Sci. Paris.* **310** (1990) 505–508
- [Bo] Borel A, Density and maximality of arithmetic groups, *J. Reine Angew. Math.* **224** (1966) 78–89
- [BH] Borel A and Harish-Chandra, Arithmetic subgroups of algebraic groups, *Ann. Math.* **75** (1962) 485–535
- [BJ] Borel A and Jacquet H, Automorphic forms and automorphic representations, *Proc. Symp. pure Math.* **33** (1979) part I, 189–202
- [C] Chernousov V, On the Hasse principle for groups of type  $E_8$ , *Sov. Math. Dokl.* **39** (1989) 592–596
- [E] Epstein P, Zur Theorie allgemeiner Zetafunktionen, I, II, *Math. Ann.* **56** (1903) 615–644; **63** (1907) 205–216
- [GJ] Godement R and Jacquet H, *Zeta functions of simple algebras*, Lect. notes in Math. No. **260**, Springer-Verlag (1972)
- [H] Harish-Chandra, *Automorphic forms on semisimple Lie groups*, Lect. notes in Math. No. **68**, Springer-Verlag (1968)
- [Hej] Hejhal D, Some Dirichlet series with coefficients related to periods of automorphic eigen forms, *Proc. Jpn Acad.* **58** (1982) 413–417

- [Hec1] Hecke E, Eine neue art von Zetafunktionen und ihre Beziehungen zur Verteilung der Primzahlen, Erste Mitteilung, *Math. Z.* 1 (1918) 357–376
- [Hec2] Hecke E, *ibid*, Zweite Mitteilung, *Math. Z.* 6 (1920) 11–51
- [Hel1] Helgason S, Fundamental solutions of invariant differential operators on a symmetric space, *Am. J. Math.* 86 (1964) 565–601
- [Hel2] Helgason S, A duality for symmetric spaces with application to group representations II, *Adv. Math.* 22 (1976) 187–219
- [Hel3] Helgason S, *Groups and geometric analysis*, Academic Press, 1984
- [Hel4] Helgason S, Some results on invariant differential operators on symmetric spaces, *Am. J. Math.* 114 (1992) 789–811
- [KS] Katok S and Sarnak P, Heegner points, cycles and Maass forms. Preprint, 1990
- [KK] Kawai T and Kashiwara M, On holonomic systems for  $\Pi_{l=1}^N (f_l + \sqrt{-10})^{k_l}$ , *Publ. RIMS, Kyoto Univ.* 15 (1979) 551–575
- [K] Kottwitz R E, On Tamagawa numbers, *Ann. Math.* 127 (1988) 629–646
- [M1] Maass H, Spherical functions and quadratic forms, *J. Ind. Math. Soc.* 20 (1956) 117–162
- [M2] Maass H, Zetafunktionen mit Größencharakteren und Kugelfunktionen, *Math. Ann.* 132 (1957) 1–32
- [M3] Maass H, Über die räumliche Verteilung der Punkte in Gittern mit indefiniter Metrik, *Math. Ann.* 138 (1959) 287–315
- [M4] Maass H, *Siegel's modular forms and Dirichlet series*, Letc. notes in Math. No. 216, Springer Verlag (1971)
- [Mat] Matsuki T, The orbits of affine symmetric spaces under the action of minimal parabolic subgroups, *J. Math. Soc. Jpn* 12 (1982) 307–320
- [Och] Ochiai H, A remark on invariant eigenfunctions on some exceptional noncompact Riemannian symmetric spaces. Preprint, 1990
- [O] Oshima T, Poisson transformations on affine symmetric spaces, *Proc. Jpn Acad.* 55 (1979) 323–327
- [OS] Oshima T and Sekiguchi J, Eigenspaces of invariant differential operators on an affine symmetric space, *Invent. Math.* 57 (1980) 1–81
- [Ro] Rossmann W, Analysis on real hyperbolic spaces, *J. Funct. Anal.* 30 (1978) 448–477
- [Ru] Rubenthaler H, *Espaces préhomogènes de type parabolique*. Thèse, Université de Strasbourg, 1982
- [Sab] Sabbah C, Proximité évanescence II, *Compos. Math.* 64 (1987) 213–241
- [S1] Sato F, Zeta functions in several variables associated with prehomogeneous vector spaces I, Functional equations, *Tôhoku Math. J.* 34 (1982) 437–483
- [S2] Sato F, *ibid*. II, A convergence criterion, *Tôhoku Math. J.* 35 (1983) 77–99
- [S3] Sato F, *ibid*. III, Eisenstein series for indefinite quadratic forms, *Ann. Math.* 116 (1982) 177–212
- [S4] Sato F, The Hamburger theorem for Epstein zeta functions, *Algebraic Analysis*, Vol. II, Academic Press (1989) 789–807
- [S5] Sato F, On functional equations of zeta distributions, *Adv. Studies in Pure Math.* 15 (1989) 465–508
- [S6] Sato F, Zeta functions with polynomial coefficients associated with prehomogeneous vector spaces. Preprint, 1989
- [S7] Sato F, The Maass zeta functions attached to positive definite quadratic forms, *Adv. Studies in pure Math.* 21 (1992) 409–443
- [S] Sato M, Theory of prehomogeneous vector spaces (Notes by T. Shintani in Japanese), *Sugaku no Ayumi* 15-1 (1970) 85–157
- [SK] Sato M and Kimura T, A classification of irreducible prehomogeneous vector spaces and their invariants, *Nagoya Math. J.* 65 (1977) 1–155
- [SS] Sato M and Shintani T, On zeta functions associated with prehomogeneous vector spaces, *Ann. Math.* 100 (1974) 131–170
- [Se] Sekiguchi J, Eigenspaces of the Laplace-Beltrami operator of a hyperboloid, *Nagoya Math. J.* 79 (1980) 151–185
- [Sh] Shintani T, On zeta functions associated with the vector space of quadratic forms, *J. Fac. Sci. Univ. Tokyo* 22 (1975) 25–65
- [Si] Siegel C, *Advanced analytic number theory*, Tata Studies in Math. 9, Tata Institute of Fundamental Research, Bombay, 1980
- [T] Terras A, *Harmonic analysis on symmetric spaces and applications II*, Springer Verlag, 1985
- [V] Vust T, Opération de groupes réductifs dans un type de cône presque homogène, *Bull. Soc. Math. France* 102 (1974) 317–334



R P BAMBAH and A C WOODS\*

Mathematics Department, Panjab University, Chandigarh 160014, India

\*Mathematics Department, Ohio State University, Columbus, Ohio 43210, USA

Dedicated to the memory of Professor K G Ramanathan

**Abstract.** A solution is given for the following Problem of G Fejes Toth: In 3-space find the thinnest lattice of balls such that every straight line meets one of the balls.

**Keywords.** Spheres (balls); lattices; thinnest arrangements.

## 1. Introduction

1.1 The object of this note is to give a solution of the following problem of G Fejes Toth [2]:

*In 3-space find the thinnest lattice arrangement of closed balls such that every straight line meets these balls.*

As pointed out by G Fejes Toth himself this is in some sense the first unsolved case of the more general problem:

*In  $n$ -space find the thinnest lattice arrangement of closed balls such that every  $k$ -dimensional ( $0 \leq k \leq n-1$ ) flat meets one of these balls.*

For  $k=0$ , this is the problem of thinnest lattice coverings by spheres, while for  $k=n-1$ , Makai [4] has shown that the problem can be reduced to that of the closest lattice packings of spheres. Thus the solution is known for  $k=0$ ,  $n \leq 5$  and for  $0 \leq k = n-1 \leq 7$ . (See any book dealing with packings and coverings, e.g. Lekkerkerker and Gruber [3]). The problem above can be generalised to one for other "bodies" also. In the case of convex bodies, Makai [4] has shown that a theorem analogous to the one for spheres holds if  $k=n-1$ . Our solution to the Fejes Toth problem stated in the beginning is contained in the following Theorems I and II and the remark after Theorem II.

(We shall throughout be working in the three-dimensional Euclidan space  $R^3$ ).

**Theorem I.** Let  $K$  be the sphere  $|x| \leq 1$ . Let  $\Lambda$  be a lattice with determinant  $d(\Lambda)$ . If every straight line meets a ball  $K + A$ ,  $A \in \Lambda$ , then  $d(\Lambda) \leq 2(4/3)^3$ .

**Theorem II.** Let  $K$  be the sphere  $|x| \leq 1$  and  $\Lambda$  be the lattice generated by  $4/3(1, 1, 0)$ ,  $4/3(0, 1, 1)$  and  $4/3(1, 0, 1)$ . Then every straight line meets a sphere  $K + A$ ,  $A \in \Lambda$ .

**Remark** Our proof of Theorem I (see §4.4) shows that "up to" orthogonal transformations the lattice  $\Lambda$  of Theorem II is the only "critical" lattice.

For convenience we replace Theorems I and II by the equivalent Theorems I', II':

**Theorem I'.** Let  $K$  be the sphere  $|x| \leq 3/4$  and  $\Lambda$  a lattice with determinant  $d(\Lambda)$ . If every straight line meets a ball  $K + A$ ,  $A \in \Lambda$ , then  $d(\Lambda) \leq 2$ .

**Theorem II'.** Let  $K$  be the sphere  $|x| \leq 3/4$  and  $\Lambda$  the lattice generated by  $(1, 1, 0)$ ,  $(0, 1, 1)$  and  $(1, 0, 1)$ . Then every straight line meets a  $K + A$ ,  $A \in \Lambda$ .

## 2. Proof of Theorem I'

2.1. Let  $K$  be the sphere  $|x| \leq 3/4$  and  $\Lambda$  a lattice. Let  $A_1 \in \Lambda$ . Let  $\Pi$  be the plane through  $O$  perpendicular to  $OA_1$ . Let  $\Lambda^*$  be the (orthogonal) projection of  $\Lambda$  on  $\Pi$ . Let  $C$  be the circle  $K \cap \Pi$ . All lines parallel to  $OA_1$  meet a  $K + A$ ,  $A \in \Lambda$  is equivalent to the statement: the circles  $C + A^*$ ,  $A^* \in \Lambda^*$  cover  $\Pi$ , i.e. the "covering radius"  $\rho(\Lambda^*)$  of  $\Lambda^*$  is  $\leq 3/4$ .

2.2. Let  $A_1, A_2, A_3$  be a basis of  $\Lambda$ . Let  $L$  be the matrix  $(A_1, A_2, A_3)$  with  $A_1, A_2, A_3$  written as column vectors. The positive definite quadratic form  $f(x) = f(x_1, x_2, x_3) = X' L' L X$ , where  $X' = (x_1, x_2, x_3)$  is called the quadratic form of  $\Lambda$  w.r.t. the basis  $A_1, A_2, A_3$ . Its determinant  $d(f) = \det(L' L) = d^2(\Lambda)$ . If  $(B_1, B_2, B_3) = (A_1, A_2, A_3)U$  is any other basis of  $\Lambda$ , the  $U \in GL(3, \mathbb{Z})$  and the corresponding quadratic form  $X' U' L' L U X$  is equivalent to  $f(X)$ . In fact the quadratic forms corresponding to different bases of  $\Lambda$  consist of the class of quadratic forms equivalent to  $f$ .

Again if  $f(x) = X' L' L X = X' M' M X$ , then  $M = TL$ , where  $T$  is orthogonal and the lattice  $T\Lambda$  with basis  $TA_1, TA_2, TA_3$  is an orthogonal transform of  $\Lambda$ . We may note that  $TK = K$ , and  $\Lambda$  has the property of Theorem I' if and only if  $T\Lambda$  has.

2.3. Let  $f(x) = \sum a_{ij} x_i x_j$ ,  $a_{ij} = a_{ji}$  be the real positive definite quadratic form corresponding to a basis  $A_1, A_2, A_3$  of  $\Lambda$ . Write

$$\begin{aligned} f &= a_{11} \left( x_1 + \frac{a_{12}}{a_{11}} x_2 + \frac{a_{13}}{a_{11}} x_3 \right)^2 + g(x_2, x_3) \\ &= (\alpha_{11} x_1 + \alpha_{12} x_2 + \alpha_{13} x_3)^2 + (\alpha_{22} x_2 + \alpha_{23} x_3)^2 + (\alpha_{32} x_2 + \alpha_{33} x_3)^2, \end{aligned}$$

and  $f$  is the quadratic form of a lattice  $\Lambda_1 = T\Lambda$ ,  $T$  orthogonal, with respect to the basis  $B_1 = TA_1, B_2 = TA_2, B_3 = TA_3$ , and  $B_1 = (\alpha_{11}, 0, 0)$ ,  $B_2 = (\alpha_{12}, \alpha_{22}, \alpha_{32})$ ,  $B_3 = (\alpha_{13}, \alpha_{23}, \alpha_{33})$ . Every line parallel to  $OA_1$  meets a  $K + A$ ,  $A \in \Lambda$  if and only if every line parallel to  $OB_1$  meets a  $K + B$ ,  $B \in \Lambda_1$ . Since  $B_1$  is the point  $(\alpha_{11}, 0, 0)$ , the plane  $\Pi$  of 2.1 is  $x_1 = 0$  and the projection  $\Lambda^*$  of  $\Lambda_1$  on  $\Pi$  is the lattice generated by  $(0, \alpha_{22}, \alpha_{32})$  and  $(0, \alpha_{23}, \alpha_{33})$ , while

$$g(x_2, x_3) = (\alpha_{22} x_2 + \alpha_{23} x_3)^2 + (\alpha_{32} x_2 + \alpha_{33} x_3)^2.$$

Let  $\rho = \rho(\Lambda^*)$  be the covering radius of  $\Lambda^*$  and  $R(g) = \rho^2$ . ( $R(g)$  depends only on  $g$ ,



because if  $g$  is a quadratic form of another lattice  $\Lambda_1^*$ , then  $\Lambda_1^* = T\Lambda^*$ , where  $T$  is orthogonal and the covering radius of  $\Lambda_1^*$  is the same as that of  $\Lambda^*$ .)

By §2.1 all lines parallel to  $OA_1$  meet a  $K + A$ ,  $A \in \Lambda$  if and only if  $\rho(\Lambda)^* \leq 3/4$ , if and only if  $R(g) \leq 9/16$ . Since every primitive lattice point can be extended to a basis of  $\Lambda$ , all lines parallel to lines  $OA$ ,  $A \in \Lambda$  meet the balls  $K + P$ ,  $P \in \Lambda$  if and only if for all forms  $f' \sim f$ , the corresponding "sections"  $g'(x_2, x_3)$  have  $R(g') \leq 9/16$ . More precisely, the hypothesis of Theorem I' implies the following:

Let  $\Lambda$  be a lattice. Let  $f(x) = \sum a_{ij}x_i x_j$ ,  $a_{ij} = a_{ji}$  be any quadratic form of  $\Lambda$ . Let

$$f(x) = a_{11} \left( x_1 + \frac{a_{12}}{a_{11}}x_2 + \frac{a_{13}}{a_{11}}x_3 \right)^2 + g(x_2, x_3).$$

Then

$$R(g) \leq 9/16.$$

To prove Theorem I' it is enough to prove

**Theorem IA.** Let  $f(x) = \sum a_{ij}x_i x_j$ ,  $a_{ij} = a_{ji}$  be a real positive definite quadratic form with determinant  $d(f)$ . Let  $f' \sim f$ ; write

$$f'(x) = a'_{11} \left( x_1 + \frac{a'_{12}}{a'_{11}}x_2 + \frac{a'_{13}}{a'_{11}}x_3 \right)^2 + g'(x_2, x_3).$$

If  $R(g') \leq 9/16$  for each  $f' \sim f$ , then  $d(f) \leq 4$ .

2.4. Let  $f(x) = \sum a_{ij}x_i x_j$ ,  $a_{ij} = a_{ji}$  be a positive definite quadratic form. Let

$$f(x) = a_{11} \left( x_1 + \frac{a_{12}}{a_{11}}x_2 + \frac{a_{13}}{a_{11}}x_3 \right)^2 + g(x_2, x_3).$$

Then

$$\begin{aligned} a_{11}g &= (a_{11}a_{22} - a_{12}^2)x_2^2 + 2(a_{11}a_{23} - a_{12}a_{13})x_2x_3 + (a_{11}a_{33} - a_{13}^2)x_3^2 \\ &= A_{33}x_2^2 - 2A_{23}x_2x_3 + A_{22}x_3^2 \\ &= G', \text{ say,} \end{aligned}$$

where  $A_{ij}$  are the entries of the matrix adjoint to  $(a_{ij})$ . Since  $g = a_{11}^{-1}G'$ ,  $R(g) = a_{11}^{-1}R(G')$ . If

$$G = A_{22}x_2^2 + 2A_{23}x_2x_3 + A_{33}x_3^2,$$

then  $G \sim G'$  and  $R(G) = R(G')$ , and

$$R(g) = a_{11}^{-1}R(G). \quad (a)$$

Let  $A = (a_{ij})$ ,  $\text{adj } A = (A_{ij})$ . Then  $A \text{ adj } A = \det(A)I$ , and  $\det(\text{adj } A) = (\det A)^2$ . Write

$$F(x) = \text{adj } f(x) = \sum A_{ij}x_i x_j$$

Then

$$d(F) = \det(A_{ij}) = (\det A)^2 = d^2(f). \quad (b)$$

Since

$$A(\text{adj } A) = (\det A)I = d(f)I, \text{ and } (\text{adj } A) \text{ adj}(\text{adj } A) = d(F)I = d^2(f)I,$$

we have

$$\frac{1}{d(f)} A = \frac{1}{d^2(f)} \text{adj}(\text{adj } A)$$

i.e.

$$\frac{1}{d(f)} (a_{ij}) = \frac{1}{d^2(f)} \text{adj}(A_{ij})$$

Equating elements in the leading position, we get

$$\begin{aligned} \frac{1}{d(f)} a_{11} &= \frac{1}{d^2(f)} (A_{22}A_{33} - A_{23}^2) \\ &= \frac{1}{d^2(f)} d(G), \end{aligned}$$

and  $a_{11}^{-1} = d(f)/d(G) = \sqrt{d(F)/d(G)}$ , and, by (a),

$$R(g) = R(G) \sqrt{d(F)/d(G)}.$$

Therefore,

$$R(g) \leq 9/16 \text{ iff } R(G) \leq 9/16 \, d(G)/d(F)^{1/2} \quad (c)$$

and

$$d(F) = d^2(f). \quad (d)$$

It is well known that if  $f \sim f'$ , then  $\text{adj } f \sim \text{adj } f'$  and vice versa, i.e., the class of forms equivalent to  $\text{adj } f$  is the class of adjoints of forms  $\sim f$ .

Let  $F(x_1, x_2, x_3) = \sum A_{ij} x_i x_j$  be a definite quadratic form and  $F_1 \sim F$ . Let  $G(x_2, x_3) = F_1(0, x_2, x_3)$  be called a partial sum of  $F$  and let  $S$  be the set of partial sums of  $F$ . Since  $F(x_1, x_2, x_3) \sim F(x_3, x_1, x_2)$  the set of partial sums of  $F$  consists of the forms  $G(x_1, x_2) = F'(x_1, x_2, 0)$  for all forms  $F' \sim F(x)$ .

We can replace Theorem IA by (see (c) and (d) above).

**Theorem IB.** Let  $F(x_1, x_2, x_3) = \sum A_{ij} x_i x_j$ ,  $A_{ij} = A_{ji}$  be a positive definite quadratic form. Suppose for every partial sum  $G$  of  $F$  we have  $R(G) \leq 9/16 \, d(G)/\sqrt{d(F)}$ . Then  $d(F) \leq 16$ .

It is clear that we can replace  $F$  by any equivalent form without affecting the hypothesis or conclusion of the theorem. For convenience we alter the notation a little bit and state Theorem IB as:

**Theorem IC.** Let  $f(x_1, x_2, x_3) = \sum a_{ij} x_i x_j$ ,  $a_{ij} = a_{ji}$  be a positive definite quadratic form. Suppose for every partial sum  $g(x_1, x_2) = f'(x_1, x_2, 0)$ , where  $f' \sim f$ , we have  $R(g) \leq 9/16 \, d(g)/\sqrt{d(f)}$ , then  $d(f) \leq 16$ .

angled with largest angle at O. In this case the covering radius of  $\Lambda$  is the circumradius of  $\Delta OAB$ . (see e.g. Dickson [1], pp. 160).

Now suppose  $A, B$  generate a two-dimensional lattice and  $\Delta OAB$  is acute angled. Then  $(A_1, B_1) = (A, B)$  or  $(-A, B - A)$  or  $(-B, A - B)$  is a reduced basis of  $\Lambda$  and its covering radius is the circumradius of  $\Delta OA_1B_1 =$  the circumradius of  $\Delta OAB$ . Thus if  $A, B$  generate  $\Lambda$  and  $\Delta OAB$  is acute angled, then the covering radius  $\rho(\Lambda)$  of  $\Lambda$  is the circumradius of  $\Delta OAB$ .

Let  $g(x, y) = ax^2 + 2bxy + cy^2$  be positive definite. Let  $g(x, y) = (\alpha x + \beta y)^2 + (\gamma x + \delta y)^2$ .

Let  $A = (x, y)$ ,  $B = (\beta, \delta)$ . Then  $A, B$  generate a lattice  $\Lambda$  and  $R(g) = \rho^2(\Lambda)$ . The triangle  $OAB$  is acute angled if the square of each side  $\leq$  sum of squares of the other two sides, i.e., if

$$a \leq c + (a + c - 2b),$$

$$c \leq a + (a + c - 2b),$$

$$a + c - 2b \leq a + c,$$

i.e.

$$b \leq c, b \leq a, b \geq 0, \text{ i.e.}$$

$$0 \leq b \leq \min(a, c).$$

Therefore, if  $0 \leq b \leq \min(a, c)$ , then

$R(g) = (\text{circumradius of triangle } OAB) = ac(a + c - 2b)/4 d(g)$ . (If  $ABC$  is an acute angle triangle with sides  $a, b, c$  circumradius  $\rho$  and area  $\Delta$ , then

$$\rho = \frac{a}{2 \sin A} = \frac{b}{2 \sin B} = \frac{c}{2 \sin C},$$

$$\begin{aligned} \rho^3 &= \frac{abc}{8 \sin A \sin B \sin C} = \frac{a^3 b^3 c^3}{64(1/2 bc \sin A)(1/2 ca \sin B)(1/2 ab \sin C)} \\ &= \frac{a^3 b^3 c^3}{64 \Delta^3} \end{aligned}$$

so that

$$\rho^2 = \frac{a^2 b^2 c^2}{4(2\Delta)^2}.$$

3.2 Let  $f(x_1) = \sum a_{ij} x_i x_j$ ,  $a_{ij} = a_{ji}$  be a positive definite form, all of whose partial sums  $g(x_1, x_2)$  have  $R(g) \leq 9/16 d(g)/\sqrt{d(f)}$ . We have to show  $d(f) \leq 16$ .

By replacing  $f$ , by an equivalent form reduced in the sense of Gauss and Sieber (see, e.g. Dickson [1], Th 103, pp. 171), we can suppose

$$0 < a_{11} \leq a_{22} \leq a_{33},$$

$$2|a_{12}| \leq a_{11}, 2|a_{13}| \leq a_{11}, 2|a_{23}| \leq a_{22}, \text{ and} \quad (A)$$

$$a_{ij}, i \neq j, \text{ all have the same sign,}$$

$$a_{11} + a_{22} + 2(a_{12} + a_{13} + a_{23}) \geq 0.$$

We divide the proof into two cases:

case I: all  $a_{ij}, i \neq j$ , are negative (or 0),

case II: all  $a_{ij}, i \neq j$ , are positive (or 0).

#### 4. Proof of Theorem IC Case I

4.1 Clearly  $g_1 = f(0, x_2, x_3)$ ,  $g_2 = f(x_1, 0, x_3)$  and  $g_3 = f(x_1, x_2, 0)$  are all partial sums of  $f$ . If  $\Sigma A_{ij} x_i x_j$  is adjoint to  $f$ , then

$$d(g_1) = A_{11}, d(g_2) = A_{22}, d(g_3) = A_{33}.$$

Also each  $g$  is equivalent to one with the cross term of opposite sign.

Therefore, applying the formula of § 3.1,

$$R(g_1) = a_{22}a_{33}(a_{22} + a_{33} + 2a_{23})/4 A_{11},$$

$$R(g_2) = a_{33}a_{11}(a_{33} + a_{11} + 2a_{31})/4 A_{22}, \text{ and}$$

$$R(g_3) = a_{11}a_{22}(a_{11} + a_{22} + 2a_{12})/4 A_{33}$$

By the hypothesis  $R(g_i) \leq 9/16 d(g_i)/\sqrt{d(f)}$ , and we have

$$a_{22}a_{33}(a_{22} + a_{33} + 2a_{23})/4 A_{11} \leq 9/16 A_{11}/\sqrt{d(f)}$$

or

$$4 a_{22}a_{33}(a_{22} + a_{33} + 2a_{23})\sqrt{d(f)} \leq 9 A_{11}^2. \quad (1)$$

Similarly,

$$4 a_{33}a_{11}(a_{33} + a_{11} + 2a_{13})\sqrt{d(f)} \leq 9 A_{22}^2, \quad (2)$$

and

$$4 a_{11}a_{22}(a_{11} + a_{22} + 2a_{12})\sqrt{d(f)} \leq 9 A_{33}^2. \quad (3)$$

4.2 Define  $\beta_{12}, \beta_{23}, \beta_{13}$  by

$$\begin{aligned} a_{12} &= -\beta_{12}\sqrt{a_{11}a_{22}}, a_{13} = -\beta_{13}\sqrt{a_{11}a_{33}}, \\ a_{23} &= -\beta_{23}\sqrt{a_{22}a_{33}}, \end{aligned} \quad (4)$$

and put

$$t_1 = (a_{11}/a_{22})^{1/2}, t_2 = (a_{22}/a_{33})^{1/2}. \quad (5)$$

The reduction conditions (A) of § 3.2 give

$$0 \leq t_1, t_2 \leq 1 \quad (6)$$

$$0 \leq \beta_{12} \leq \frac{1}{2}t_1, 0 \leq \beta_{13} \leq \frac{1}{2}t_1t_2, 0 \leq \beta_{23} \leq \frac{1}{2}t_2, \quad (7)$$

and

$a_{11} + a_{22} + 2(a_{12} + a_{13} + a_{23}) \geq 0$  becomes

$$a_{11} + a_{22} \geq 2(\beta_{12}\sqrt{a_{11}a_{22}} + \beta_{13}\sqrt{a_{11}a_{33}} + \beta_{23}\sqrt{a_{22}a_{33}}),$$

so that, dividing by  $\sqrt{a_{22}a_{33}}$ , we get

$$t_1^2 t_2 + t_2 \geq 2(\beta_{12} t_1 t_2 + \beta_{13} t_1 + \beta_{23}). \quad (8)$$

Now, if we write

$$g(t_1, t_2) = t_1^2 t_2 + t_2 - 2(\beta_{12} t_1 t_2 + \beta_{13} t_1 + \beta_{23}),$$

then

$$\frac{\partial g}{\partial t_1} = 2t_1 t_2 - 2\beta_{12} t_2 - 2\beta_{13} \geq 2t_1 t_2 - t_1 t_2 - t_1 t_2 \quad (\text{By (7)})$$

$$\geq 0,$$

$$\begin{aligned} \frac{\partial g}{\partial t_2} &= t_1^2 + 1 - 2\beta_{12} t_1 \\ &= 1 + t_1(t_1 - 2\beta_{12}) \geq 1 > 0. \quad (\text{By (7)}) \end{aligned}$$

Therefore, (8) remains true if we replace  $t_1, t_2$  by 1, i.e.

$$\beta_{12} + \beta_{13} + \beta_{23} \leq 1. \quad (B)$$

Also,

$$\begin{aligned} d(f) &= a_{11} a_{22} a_{33} + 2a_{12} a_{13} a_{23} - a_{33} a_{12}^2 - a_{11} a_{23}^2 - a_{12} a_{13}^2 \\ &= a_{11} a_{22} a_{33} (1 - 2\beta_{12} \beta_{13} \beta_{23} - \beta_{12}^2 - \beta_{13}^2 - \beta_{23}^2) \\ &= a_{11} a_{22} a_{23} \Delta, \text{ say.} \end{aligned} \quad (C)$$

4.3 Using inequality 1 of § 4.1, together with the arithmetic geometric mean inequality, we get

$$\begin{aligned} 9A_{11}^2 &\geq 4a_{22} a_{33} (a_{22} + a_{33} + 2a_{23}) \sqrt{d(f)} \\ &\geq 8a_{22} a_{33} (\sqrt{a_{22} a_{33}} + a_{23}) \sqrt{d(f)} \\ &= 8a_{22} a_{33} \sqrt{a_{11} a_{22} a_{33} \Delta} (\sqrt{a_{22} a_{33}} + a_{23}) \\ &= 8\sqrt{a_{11} \Delta} (a_{22} a_{33})^{3/2} (\sqrt{a_{22} a_{33}} + a_{23}), \end{aligned}$$

so that

$$\begin{aligned} 8\sqrt{a_{11} \Delta} &\leq 9(a_{22} a_{33} - a_{23}^2)^2 / (a_{22} a_{33})^{3/2} (\sqrt{a_{22} a_{33}} + a_{23}) \\ &= 9 \left\{ 1 - \frac{a_{23}^2}{a_{22} a_{33}} \right\}^2 / \left\{ 1 + \frac{a_{23}}{\sqrt{a_{22} a_{33}}} \right\} \\ &= 9(1 - \beta_{23}^2)^2 / (1 - \beta_{23}) \\ &= 9(1 - \beta_{23})(1 + \beta_{23})^2, \text{ and} \\ \sqrt{a_{11} \Delta} &\leq \frac{9}{8} (1 - \beta_{23})(1 + \beta_{23})^2. \end{aligned} \quad (9)$$

Similarly, (2), (3) give

$$\sqrt{a_{22} \Delta} \leq \frac{9}{8} (1 - \beta_{31})(1 + \beta_{31})^2, \quad (10)$$

and

$$\sqrt{a_{33}}\Delta \leq \frac{9}{8}(1 - \beta_{12})(1 + \beta_{12})^2 \quad (11)$$

Multiplying (9), (10), and (11), we get

$$\begin{aligned} \sqrt{d(f)} &= \sqrt{a_{11}a_{22}a_{33}}\Delta \leq (9/8)^3(1 - \beta_{12})(1 - \beta_{23})(1 - \beta_{13}) \\ &\quad (1 + \beta_{12})^2(1 + \beta_{23})^2(1 + \beta_{13})^2/\Delta \\ &= h(\beta_{12}, \beta_{23}, \beta_{13}), \text{ say} \end{aligned} \quad (D)$$

4.4 Our object now is to use (D) above to show that the condition (B) of §4.2 (i.e.  $\beta_{12} + \beta_{23} + \beta_{13} \leq 1$ ) implies  $\sqrt{d(f)} \leq 4$ . (This will, of course, prove theorem IC in case I).

We note that if  $\beta_{12} + \beta_{23} + \beta_{13} \leq 1$ , one of the  $\beta$ 's must be  $\leq 1/3$ . Increasing the  $\beta$  increases the numerator of  $h$  and decreases its denominator

$$\Delta = (1 - 2\beta_{12}\beta_{23}\beta_{13} - \beta_{12}^2 - \beta_{13}^2 - \beta_{23}^2),$$

because

$$\begin{aligned} \frac{d}{dx}(1 - x)(1 + x)^2 &= -(1 + x)^2 + 2(1 - x^2) \\ &= (1 + x)(1 - 3x) \geq 0 \text{ if } x \leq 1/3. \end{aligned}$$

Increasing the  $\beta$ 's appropriately, we can assume

$$\beta_{12} + \beta_{23} + \beta_{13} = 1. \quad (E)$$

Putting  $\beta_{23} = 1 - \beta_{12} - \beta_{13}$ , we have

$$\begin{aligned} \Delta &= 1 - 2\beta_{12}\beta_{13}\beta_{23} - \beta_{12}^2 - \beta_{13}^2 - \beta_{23}^2 \\ &= 1 - 2\beta_{12}\beta_{13}(1 - \beta_{12} - \beta_{13}) - \beta_{12}^2 - \beta_{13}^2 - (1 - \beta_{12} - \beta_{13})^2 \\ &= 1 - 2\beta_{12}\beta_{13} + 2\beta_{12}\beta_{13}(\beta_{12} + \beta_{13}) - \beta_{12}^2 - \beta_{13}^2 \\ &\quad - 1 + 2(\beta_{12} + \beta_{13}) - (\beta_{12} + \beta_{13})^2 \\ &= 2(\beta_{12} + \beta_{13})(1 + \beta_{12}\beta_{13} - \beta_{12} - \beta_{13}) \\ &= 2(\beta_{12} + \beta_{13})(1 - \beta_{12})(1 - \beta_{13}), \end{aligned} \quad (12)$$

while

$$\begin{aligned} (1 - \beta_{12})(1 - \beta_{13})(1 - \beta_{23})(1 + \beta_{12})^2(1 + \beta_{13})^2(1 + \beta_{23})^2 \\ = (1 - \beta_{12})(1 - \beta_{13})(\beta_{12} + \beta_{13})(1 + \beta_{12})^2(1 + \beta_{13})^2(2 - \beta_{12} - \beta_{13})^2, \end{aligned}$$

so that (D) gives

$$\begin{aligned} \sqrt{d(f)} &\leq (9/8)^3(1 - \beta_{12})(1 - \beta_{13})(\beta_{12} + \beta_{13})(1 + \beta_{12})^2(1 + \beta_{13})^2 \\ &\quad (2 - \beta_{12} - \beta_{13})^2/2(\beta_{12} + \beta_{13})(1 - \beta_{12})(1 - \beta_{13}) \\ &= (9^3/2^{10})(1 + \beta_{12}^2)(1 + \beta_{13}^2)(2 - \beta_{12} - \beta_{13})^2. \end{aligned} \quad (F)$$

Also (7) gives  $0 \leq \beta_{12} \leq 1/2$ ,  $0 \leq \beta_{13} \leq 1/2$ . We now observe

*Lemma.* The maximum of  $f(x, y) = (1+x)(1+y)(2-x-y)$ , subject to  $0 \leq x, y \leq 1$  is attained only when  $x = y = 1/3$  and has the value  $(4/3)^3$ .

*Proof.* By the inequality of arithmetic geometric mean

$$f(x, y) = (1+x)(1+y)(2-x-y) \leq \left( \frac{1+x+1+y+2-x-y}{3} \right)^3 = (4/3)^3,$$

and the equality occurs if  $1+x = 1+y = 2-x-y = 4/3$ , i.e.  $x = y = 1/3$ .

Using the Lemma in (F), we get

$$\sqrt{d(f)} \leq \frac{9^3}{2^{10}} (4/3)^6 = 2^2 = 4,$$

which proves Theorem I(C) in this case.

We also note that  $d(f)$  can be 16 only if

$$\beta_{12} = 1/3, \beta_{13} = 1/3, \beta_{23} = 1/3,$$

$$\Delta = 2 \frac{2}{3} \frac{2}{3} \frac{2}{3} = 2(2/3)^3,$$

and by (9), (10), (11)

$$\sqrt{a_{ii}\Delta} = \frac{9}{8} 2/3 (4/3)^2$$

i.e.

$$a_{ii} = (4/3)^2 \frac{3^3}{16} = 3,$$

i.e.,

$$f(x_1, x_2, x_3) = 3 \sum_{1 \leq i \leq 3} x_i^2 - 2 \sum_{1 \leq i < j \leq 3} x_i x_j$$

## 5. Proof of Theorem IC, Case II

5.1 In this case  $f = \sum a_{ij} x_i x_j$ ,  $a_{ij} = a_{ji}$ ; and

$$0 < a_{11} \leq a_{22} \leq a_{33},$$

$$0 \leq 2a_{12}, 2a_{13} \leq a_{11}, 0 \leq 2a_{23} \leq a_{22}.$$

Writing

$$a_{ij} = \beta_{ij} \sqrt{a_{ii} a_{jj}}, \quad i \neq j$$

We have

$$0 \leq \beta_{ij} \leq \frac{1}{2}.$$

## 6. Proof of Theorem IC Case II (a)

6.1 As in § 4.1, considering the partial sums  $f(0, x_2, x_3)$ ,  $f(x_1, 0, x_3)$ ,  $f(x_1, x_2, 0)$ , and noting  $a_{ij} \geq 0$ , we get

$$4a_{22}a_{33}(a_{22} + a_{33} - 2a_{23})\sqrt{d(f)} \leq 9A_{11}^2, \quad (1')$$

$$4a_{33}a_{11}(a_{33} + a_{11} - 2a_{13})\sqrt{d(f)} \leq 9A_{22}^2, \quad (2')$$

and

$$4a_{11}a_{22}(a_{11} + a_{22} - 2a_{12})\sqrt{d(f)} \leq 9A_{33}^2 \quad (3')$$

Also

$$\begin{aligned} d(f) &= a_{11}a_{22}a_{33} + 2a_{12}a_{13}a_{23} - a_{11}a_{23}^2 - a_{22}a_{13}^2 - a_{33}a_{12}^2 \\ &= a_{11}a_{22}a_{33}(1 + 2\beta_{12}\beta_{22}\beta_{33} - \beta_{12}^2 - \beta_{23}^2 - \beta_{31}^2) \\ &= a_{11}a_{22}a_{33}\Delta', \text{ say} \end{aligned} \quad (C')$$

from (1') and (C') we get, applying A-G mean inequality,

$$\begin{aligned} 9A_{11}^2 &\geq 8a_{22}a_{23}(\sqrt{a_{22}a_{33}} - a_{23})\sqrt{d(f)} \\ &= 8\sqrt{a_{11}\Delta'}(a_{22}a_{33})^2(1 - \beta_{23}), \end{aligned}$$

so that

$$\begin{aligned} 8\sqrt{a_{11}\Delta'} &\leq 9(a_{22}a_{33} - a_{23}^2)/(a_{22}a_{33})^2(1 - \beta_{23}) \\ &= 9(1 - \beta_{23}^2)^2/(1 - \beta_{23}) \\ &= 9(1 - \beta_{23})(1 + \beta_{23})^2 \end{aligned} \quad (4')$$

Similarly, (2'), (3') and (C') give

$$8\sqrt{a_{22}\Delta'} \leq 9(1 - \beta_{13})(1 + \beta_{13})^2 \quad (5')$$

$$8\sqrt{a_{33}\Delta'} \leq 9(1 - \beta_{12})(1 + \beta_{12})^2. \quad (6')$$

Multiplying (4'), (5'), (6'), we get

$$\begin{aligned} 8^3\sqrt{d(f)}\Delta' &\leq 9^3(1 - \beta_{12})(1 - \beta_{13})(1 - \beta_{23}) \\ &\quad (1 + \beta_{12})^2(1 + \beta_{13})^2(1 + \beta_{23})^2, \end{aligned}$$

and

$$\begin{aligned} \sqrt{d(f)} &\leq (9/8)^3(1 - \beta_{12})(1 - \beta_{13})(1 - \beta_{23}) \\ &\quad (1 + \beta_{12})^2(1 + \beta_{13})^2(1 + \beta_{23})^2/1 + 2\beta_{12}\beta_{13}\beta_{23} - \beta_{12}^2 - \beta_{13}^2 - \beta_{23}^2 \\ &= F, \text{ say.} \end{aligned} \quad (F')$$



Make the substitution

$$x_1 = 1 + \beta_{12}, x_2 = 1 + \beta_{13}, x_3 = 1 + \beta_{23}.$$

Then

$$1 \leq x_i \leq 3/2, \text{ and at least one } x_i \leq 1.459.$$

Noting

$$\begin{aligned} & 2x_1x_2x_3 - (x_1 + x_2 + x_3 - 2)^2 \\ &= 2(1 + \beta_{12})(1 + \beta_{13})(1 + \beta_{23}) - (1 + \beta_{12} + \beta_{13} + \beta_{23})^2 \\ &= 1 + 2\beta_{12}\beta_{13}\beta_{23} - \beta_{12}^2 - \beta_{13}^2 - \beta_{23}^2 = \Delta', \end{aligned}$$

We get, from (F'),

$$\begin{aligned} \sqrt{d(f)} &\leq (9/8)^3(2 - x_1)(2 - x_2)(2 - x_3)x_1^2x_2^2x_3^2/ \\ &\quad 2x_1x_2x_3 - (x_1 + x_2 + x_3 - 2)^2. \\ &= F(x_1, x_2, x_3), \text{ say.} \end{aligned}$$

It is, therefore, enough to prove that if  $1 \leq x_i \leq 3/2$  and at least one  $x_i \leq 1.459$ , then  $F(x_1, x_2, x_3) \leq 4$ .

Now  $\partial F/\partial x_1$  has the same sign as

$$\begin{aligned} & (4x_1 - 3x_1^2)(2x_1x_2x_3 - (x_1 + x_2 + x_3 - 2)^2) \\ & - (2x_2x_3 - 2(x_1 + x_2 + x_3 - 2))x_1^2(2 - x_1), \end{aligned}$$

which has the same sign as

$$\begin{aligned} & (4 - 3x_1)(2x_1x_2x_3 - (x_1 + x_2 + x_3 - 2)^2) \\ & - 2x_1(2 - x_1)(x_2x_3 - x_1 - x_2 - x_3 + 2) \\ &= 4x_1x_2x_3(1 - x_1) + (x_1 + x_2 + x_3 - 2) \\ &\quad \{4x_1 - 2x_1^2 - (4 - 3x_1)(x_1 + x_2 + x_3 - 2)\} \\ &= 4x_1x_2x_3(1 - x_1) + (x_1 + x_2 + x_3 - 2) \\ &\quad \{x_1^2 - (4 - 3x_1)(x_1 + x_3 - 2)\} \\ &= G(x_1, x_2, x_3), \text{ say.} \end{aligned}$$

Writing  $x = ((x_2 + x_3)/2)$ , and noting,

$$\begin{aligned} x_2x_3 &\leq ((x_2 + x_3)/2)^2 = x^2, \quad 1 - x_1 \leq 0, \\ G(x_1, x_2, x_3) &\geq 4x_1x^2(1 - x_1) + (x_1 + 2x - 2) \\ &\quad \{x_1^2 - (4 - 3x_1)(2x - 2)\} \\ &= (x_1 - 2)^2 \{x_1 - 4(x - 1)^2\} \\ &= (x_1 - 2)^2 \{x_1 - 1 + 1 - 4(x - 1)^2\} \\ &\geq (x_1 - 2)^2(x_1 - 1), \quad (\text{because } 0 \leq x - 1 \leq \frac{1}{2}) \\ &\geq 0. \end{aligned}$$

Therefore,  $(\partial F/\partial x_1) \geq 0$ . Similarly  $(\partial F/\partial x_2) \geq 0$ ,  $(\partial F/\partial x_3) \geq 0$ , and the maximum of  $F$  will occur at  $x_1 = 1.459$ ,  $x_2 = 1.5$ ,  $x_3 = 1.5$ , so that  $F \leq F(1.459, 1.5, 1.5) = 3.99... < 4$ , and the Theorem is proved in this case.

## 7. Proof of Theorem IC Case II (b)

7.1 In this case  $0.459 \leq \beta_{ij} \leq 0.5$  for all  $i, j$ ,  $i \neq j$ . We first note that the inequality (1'), (2'), (3') of § 6.1 is valid in this case also.

Since

$$f(x_1, x_2, x_3) \sim f(x_1 - x_2, x_2, x_3),$$

The form

$$g(x_2, x_3) = f(-x_2, x_2, x_3) = (a_{11} + a_{22} - 2a_{12})x_2^2 + 2(a_{23} - a_{13})x_2x_3 + a_{33}x_3^2$$

is a partial sum of  $f$ .

Since

$$g(x_2, x_3) \sim g(x_2, -x_3),$$

$$g(x_2, x_3) \sim (a_{11} + a_{22} - 2a_{12})x_2^2 - 2|a_{23} - a_{13}|x_2x_3 + a_{33}x_3^2 = g'(x_2x_3), \text{ say.}$$

Then  $R(g) = R(g')$ .

Since

$$0 \leq 2|a_{23} - a_{13}| \leq \max(2a_{23}, 2a_{13}) \leq a_{22} \leq a_{22} + a_{11} - 2a_{12},$$

and

$$|2(a_{23} - a_{13})| \leq a_{22} \leq a_{33},$$

$$R(g) = R(g') = a_{33}(a_{11} + a_{22} - 2a_{12})$$

$$(a_{11} + a_{22} - 2a_{12} + a_{33} - 2|a_{23} - a_{13}|)/4d(g),$$

where

$$\begin{aligned} d(g) &= (a_{11} + a_{22} - 2a_{12})a_{33} - (a_{23} - a_{13})^2 \\ &= A_{11} + A_{22} + 2(a_{23}a_{13} - a_{12}a_{33}) \\ &= A_{11} + A_{22} + 2A_{12}. \end{aligned}$$

Since

$$R(g) \leq \frac{9}{16}d(g)/\sqrt{d(f)},$$

we have

$$\begin{aligned} &a_{33}(a_{11} + a_{22} - 2a_{12})(a_{11} + a_{22} + a_{33} - 2a_{12} - 2|a_{23} - a_{13}|) \\ &\quad \sqrt{d(f)} \leq \frac{9}{4}(A_{11} + A_{22} + 2A_{12})^2. \end{aligned} \quad (13)$$

Permuting  $x_1, x_2, x_3$ , we get two similar inequalities.

Using

$$\begin{aligned} \beta_{ij}(a_{ii}a_{jj})^{1/2} &= a_{ij}, \quad t_1 = \sqrt{a_{11}/a_{22}}, \quad t_2 = \sqrt{a_{22}/a_{33}}, \text{ we have} \\ (a_{11} + a_{22} - 2a_{12}) &= (a_{11}a_{22})^{1/2}(t_1 + t_1^{-1} - 2\beta_{12}), \end{aligned}$$

$$\begin{aligned}
& (a_{11} + a_{22} + a_{33} - 2a_1a_2 - 2|a_{23} - a_{13}|) \\
& = (a_{11}a_{22})^{1/2} \left\{ t_1 + t_1^{-1} + \frac{1}{t_1t_2^2} - 2\beta_{12} - \frac{2}{t_1t_2} |\beta_{23} - \beta_{13}t_1| \right\}, \\
& A_{11} + A_{22} + 2A_{12} = a_{22}a_{33} - a_{23}^2 + a_{11}a_{33} - a_{22}^2 \\
& \quad + 2(a_{23}a_{13} - a_{12}a_{33}) \\
& = a_{22}a_{33}(1 - \beta_{23}^2) + a_{11}a_{33}(1 - \beta_{13}^2) \\
& \quad + 2(a_{11}a_{22})^{1/2}(\beta_{23}\beta_{13}a_{33} - \beta_{12}a_{33}) \\
& = a_{33}(a_{11}a_{22})^{1/2} \{ t_1^{-1}(1 - \beta_{23}^2) + t_1(1 - \beta_{13}^2) + 2(\beta_{13}\beta_{23} - \beta_{12}) \},
\end{aligned}$$

and (13) becomes

$$\begin{aligned}
\sqrt{d(f)} & \leq \frac{9}{4} a_{33}^2 a_{11} a_{22} [t_1(1 - \beta_{13}^2) + t_1^{-1}(1 - \beta_{23}^2) \\
& \quad + 2(\beta_{13}\beta_{23} - \beta_{12})] / a_{33} a_{11} a_{22} (t_1 + t_1^{-1} - 2\beta_{12}) \left( t_1 + t_1^{-1} + \frac{1}{t_1t_2^2} \right. \\
& \quad \left. - 2\beta_{12} - \frac{2}{t_1t_2} |\beta_{23}\beta_{13}t_1| \right)
\end{aligned}$$

or

$$\begin{aligned}
\sqrt{d(f)} & \leq \frac{9}{4} a_{33} [t_1(1 - \beta_{13}^2) + t_1^{-1}(1 - \beta_{23}^2) \\
& \quad + 2(\beta_{13}\beta_{23} - \beta_{12})]^2 / (t_1 + t_1^{-1} - 2\beta_{12}) \\
& \quad \left( t_1 + t_1^{-1} + \frac{1}{t_1t_2^2} - 2\beta_{12} - \frac{2}{t_1t_2} |\beta_{23} - \beta_{13}t_1| \right). \tag{14}
\end{aligned}$$

Now (3') can be written as

$$\begin{aligned}
\sqrt{d(f)} & \leq \frac{9}{4} A_{33}^2 / a_{11} a_{22} (a_{11} + a_{22} - 2a_{12}) \\
& = \frac{9}{4} (a_{11}a_{22} - a_{12}^2)^2 / a_{11} a_{22} (a_{11} + a_{22} - 2a_{12}) \\
& = \frac{9}{4} (a_{11}a_{22})^{1/2} (1 - \beta_{12}^2)^2 / (t_1 + t_1^{-1} - 2\beta_{12}),
\end{aligned}$$

using (C'), we have

$$\sqrt{a_{11}a_{22}a_{33}\Delta'} \leq \frac{9}{4} (a_{11}a_{22})^{1/2} (1 - \beta_{12}^2)^2 / (t_1 + t_1^{-1} - 2\beta_{12}),$$

so that

$$a_{33} \leq (9/4)^2 (1 - \beta_{12}^2)^4 / (t_1 + t_1^{-1} - 2\beta_{12})^2 \Delta'.$$

Substituting in (14), we get

$$\begin{aligned} \sqrt{d(f)} &\leq (9/4)^3 (1 - \beta_{12}^2)^4 [t_1 (1 - \beta_{13}^2) \\ &\quad + t_1^{-1} (1 - \beta_{23}^2) + 2(\beta_{13}\beta_{23} - \beta_{12})]^2 / \\ &\quad \Delta' (t_1 + t_1^{-1} - 2\beta_{12})^3 \left( t_1 + t_1^{-1} + \frac{1}{t_1 t_2^2} - 2\beta_{12} - \frac{2}{t_1 t_2} \right. \\ &\quad \left. |\beta_{23} - \beta_{13} t_1| \right). \end{aligned} \quad (15)$$

Since

$$t_1 \leq 1; |\beta_{23} - \beta_{13} t_1| \leq 1, 1/t_2 \geq 1,$$

$$2x - 2|\beta_{23} - \beta_{13} t_1| \geq 0 \text{ if } x = \frac{1}{t_2} \geq 1,$$

$$\begin{aligned} \sqrt{d(f)} &\leq (9/4)^3 (1 - \beta_{12}^2)^4 [t_1 (1 - \beta_{13}^2) + t_1^{-1} (1 - \beta_{23}^2) \\ &\quad + 2(\beta_{13}\beta_{23} - \beta_{12})]^2 / \Delta' \{ (t_1 + t_1^{-1} - 2\beta_{12} - 2t_1^{-1} \\ &\quad |\beta_{23} - \beta_{13} t_1| (t_1 + t_1^{-1} - 2\beta_{12})^3 \}. \end{aligned} \quad (16)$$

Writing  $t$  for  $t_1$  for convenience, we have

$$0 \leq t \leq 1,$$

and

$$\begin{aligned} \sqrt{d(f)} &\leq (9/4)^3 (1 - \beta_{12}^2)^4 [t + t^{-1} - 2\beta_{12} - t\beta_{13}^2 - t^{-1}\beta_{23}^2 + 2\beta_{13}\beta_{23}]^2 / \\ &\quad \Delta' (t + t^{-1} - 2\beta_{12})^3 (t + 2t^{-1} - 2\beta_{12} - 2t^{-1}|\beta_{23} - \beta_{13}t|). \end{aligned} \quad (17)$$

Since

$$0 \leq \beta_{13}, \beta_{23}, \beta_{12} \leq \frac{1}{2},$$

$$t(1 - \beta_{13}^2) + t^{-1}(1 - \beta_{23}^2) + 2(\beta_{13}\beta_{23} - \beta_{12})$$

$$\geq 3/4(t + t^{-1}) - 2\beta_{12}$$

$$\geq 3/2 - 1 > 0,$$

$$2\beta_{13}\beta_{23} \leq t\beta_{13}^2 + t^{-1}\beta_{23}^2,$$

we have, from (17),

$$\begin{aligned} \sqrt{d(f)} &\leq (9/4)^3 (1 - \beta_{12}^2)^4 (t + t^{-1} - 2\beta_{12})^2 / \\ &\quad \Delta' (t + t^{-1} - 2\beta_{12})^3 (t + 2t^{-1} - 2\beta_{12} - 2t^{-1}|\beta_{23} - \beta_{13}t|) \\ &= (9/4)^3 (1 - \beta_{12}^2)^4 / \\ &\quad \Delta' (t + t^{-1} - 2\beta_{12}) (t + 2t^{-1} - 2\beta_{12} - 2t^{-1}|\beta_{23} - \beta_{13}t|). \end{aligned} \quad (18)$$

Now, let

$$F(t) = t + \frac{2}{t} - 2\beta_{12} - \frac{2}{t}|\beta_{23} - \beta_{13}t|.$$

If  $\beta_{23} \geq \beta_{13}t$ ,

$$\begin{aligned} F'(t) &= 1 - \frac{2}{t^2} + \frac{2\beta_{23}}{t^2} \leq 1 - \frac{2}{t^2} + \frac{1}{t^2} \\ &= 1 - \frac{1}{t^2} \leq 0, \text{ because } t \leq 1, \end{aligned}$$

while, if  $\beta_{23} < \beta_{13}t$ ,

$$F'(t) = 1 - \frac{2}{t^2} - \frac{2\beta_{23}}{t^2} \leq 1 - \frac{2}{t^2} < 0.$$

Therefore, in all cases,

$$\begin{aligned} F(t) &\geq F(1) \\ &= 3 - 2\beta_{12} - 2|\beta_{23} - \beta_{13}| \\ &\geq 3 - 2\beta_{12} - 2 \times 0.041 \\ &= 2.918 - 2\beta_{12}, \end{aligned}$$

because  $|\beta_{23} - \beta_{13}| \leq 0.5 - 0.459 = 0.041$ .

Also

$$t + \frac{1}{t} - 2\beta_{12} \geq 2 - 2\beta_{12}.$$

Therefore, (18) implies

$$\sqrt{d(f)} \leq (9/4)^3 (1 - \beta_{12}^2)^4 / (2.918 - 2\beta_{12})(2 - 2\beta_{12}) \Delta'. \quad (19)$$

Now

$$\begin{aligned} \Delta' &= 1 + 2\beta_{12}\beta_{13}\beta_{23} - \beta_{12}^2 - \beta_{13}^2 - \beta_{23}^2, \\ \frac{\partial \Delta'}{\partial \beta_{13}} &= 2\beta_{12}\beta_{23} - 2\beta_{13} \\ &\leq 2\frac{11}{22} - 2(0.459) \\ &< 0. \end{aligned}$$

Similarly

$$\frac{\partial \Delta'}{\partial \beta_{23}} < 0,$$

therefore,

$$\begin{aligned} \Delta' &\geq 1 + 2\beta_{12}\frac{11}{22} - \beta_{12}^2 - \frac{1}{4} - \frac{1}{4} \\ &= \frac{1}{2}(1 + \beta_{12} - 2\beta_{12}^2) \\ &= \frac{1}{2}(1 - \beta_{12})(1 + 2\beta_{12}). \end{aligned} \quad (20)$$

Writing  $\beta$  for  $\beta_{12}$ , for convenience, (19) gives

$$\begin{aligned}\sqrt{d(f)} &\leq (9/4)^3 (1 - \beta^2)^4 / (2 \cdot 918 - 2\beta)(1 + 2\beta)(1 - \beta)^2 \\ &= \frac{1}{2} (9/4)^3 \frac{(1 - \beta)^2 (1 + \beta)^4}{(1 \cdot 459 - \beta)(1 + 2\beta)} \\ &= \frac{1}{2} (9/4)^3 \frac{1 - \beta}{1 \cdot 459 - \beta} \frac{(1 - \beta)(1 + \beta)^4}{1 + 2\beta} \\ &= \frac{1}{2} (9/4)^3 g(\beta) h(\beta), \text{ say.}\end{aligned}\tag{21}$$

Now

$$g(\beta) = \frac{1 - \beta}{1 \cdot 459 - \beta} = 1 - \frac{0 \cdot 459}{1 \cdot 459 - \beta}$$

is a decreasing function of  $\beta$ . Therefore,

$$g(\beta) \leq g(0 \cdot 459).\tag{22}$$

Again

$$\begin{aligned}h(\beta) &= (1 - \beta)(1 + \beta)^4 / (1 + 2\beta), \\ \frac{h'(\beta)}{h(\beta)} &= \frac{-1}{1 - \beta} + \frac{4}{1 + \beta} - \frac{2}{1 + 2\beta} \\ &= \frac{4 + 4\beta - 8\beta^2 - 1 - 3\beta - 2\beta^2 - 2 + 2\beta^2}{(1 - \beta)^2 (1 + 2\beta)} \\ &= -\frac{(8\beta^2 - \beta - 1)}{(1 - \beta)^2 (1 + 2\beta)} < 0,\end{aligned}$$

because

$$\begin{aligned}8\beta^2 - \beta - 1 &\geq 8(0 \cdot 459)^2 - (0 \cdot 459) - 1 \\ &> 8(0 \cdot 459)^2 - (0 \cdot 459) - 1 \\ &= 1 \cdot 62 - 1 \cdot 459 > 0.\end{aligned}$$

Therefore

$$\begin{aligned}h(\beta) &\leq h(0 \cdot 459), \text{ and} \\ \sqrt{d(f)} &\leq \frac{1}{2} (9/4)^3 g(0 \cdot 459) h(0 \cdot 459) \\ &= \frac{729 (1 - 0 \cdot 459)^2 (1 + 0 \cdot 459)^4}{128 \cdot 1 (1 + 0 \cdot 918)} = 3 \cdot 93 \dots < 4.\end{aligned}$$

Thus  $d(f) < 16$  in this case also and the proof of Theorem IC is complete.

## 8. Proof of Theorem II'

8.1 Let  $K$  be the sphere  $|x| \leq 3/4$  and  $\Lambda$  the lattice generated by  $(1, 1, 0)$ ,  $(0, 1, 1)$ ,  $(1, 0, 1)$ . We have to show that every straight line  $l$  meets a  $k + A$ ,  $A \in \Lambda$ .

We divide the proof into two parts:

- (a) The lines  $l$  are parallel to "lattice lines"  $OA$ ,  $A \in \Lambda$ ,  
 (b)  $l$  is not parallel to any lattice line.

## 9. Proof of Theorem II' Case (a)

### 9.1 The quadratic form

$$\begin{aligned} f(x_1, x_2, x_3) &= (x_1 + x_2)^2 + (x_2 + x_3)^2 + (x_3 + x_1)^2 \\ &= 2\sum x_i^2 + 2 \sum_{1 \leq i < j \leq 3} x_i x_j \end{aligned}$$

is the quadratic form of  $\Lambda$  corresponding to the given basis. The adjoint of  $f$  is

$$F(x_1, x_2, x_3) = 3\sum x_i^2 - 2 \sum_{1 \leq i < j \leq 3} x_i x_j.$$

As explained in §2.3, Theorem II' in case (a) will follow if we can show that for every partial sum  $G$  of  $F$ ,  $R(G) \leq \frac{9}{16} d(G) / \sqrt{d(F)}$ . We note that  $F(x_1, x_2, x_3) = (x_1 + x_2 - x_3)^2 + (x_2 + x_3 - x_1)^2 + (x_3 + x_1 - x_2)^2$ . For integers  $x_i$ ,  $x_1 + x_2 - x_3$ ,  $x_2 + x_3 - x_1$ ,  $x_3 + x_1 - x_2$  are all even or all odd. Therefore, the possible non-zero values of  $F$  for integers  $x_i$  are 3, 4, 8, 11, ... in ascending order, i.e. the values can be 3, 4 or  $\geq 8$ .

Let  $G'(x_1, x_2)$  be a partial sum of  $F$  and  $G(x_1, x_2) = ax_1^2 + 2bx_1x_2 + cx_2^2$ ,  $0 \leq 2b \leq a \leq c$ ,  $a > 0$ , be the reduced form equivalent to  $G'$ . Then

$$R(G') = R(G) = ac(a + c - 2b) / 4(ac - b^2)$$

and we have to prove

$$ac(a + c - 2b) \leq 9/16(ac - b^2)^2, \quad (\text{I})$$

because  $d(F) = 16$ .

We shall prove this by contradiction, i.e. we shall show that

$$ac(a + c - 2b) > 9/16(ac - b^2)^2$$

is not possible.

Since the values of  $G$  for integers  $x_i$  are a subset of the values of  $F$  for integers  $x_i$ , we have the following possibilities:

- (i)  $a = 3$ , (ii)  $a = 4$ , (iii)  $a \geq 8$ .  
 (i)  $a = 3$ , so that  $b = 0$  or  $1$ ,  $c \geq 3$ .

If  $b = 0$ ,  $ac(a + c - 2b) > 9/16(ac - b^2)^2$ , then

$$3ac(3 + c) > 9/16(3c)^2$$

i.e.

$$11c^2 - 48c < 0$$

i.e.

$$c(11c - 48) < 0,$$

and

$$c = 3 \text{ or } c = 4, \text{ and}$$

$$G(x_1, x_2) = 3x_1^2 + 3x_2^2 \text{ or } 3x_1^2 + 4x_2^2$$

takes the value 6 or 7 for integers  $x_i$ . Since 6, 7 are not possible values of  $F$ , this case is not possible. If

$$b = 1, ac(a + c - 2b) > 9/16(ac - b^2)^2,$$

then

$$16c(1 + c) > 3(3c - 1)^2$$

i.e.

$$11c^2 - 34c + 3 < 0$$

i.e.

$$(c - 3)(11c - 1) < 0,$$

which is impossible, because  $c \geq 3$ .

(ii) Let  $a = 4$ , so that  $b = 0, 1$  or  $2$  and  $c \geq 4$ .

Then  $ac(a + c - 2b) > 9/16(ac - b^2)^2$  implies

$$64c(4 + c - 2b) > 9(4c - b^2)^2$$

or

$$80c^2 - c(72b^2 - 128b + 256) + 9b^4 < 0.$$

$b = 0$  gives

$$80c^2 < 256c$$

and  $c < 4$ , which is impossible,

$b = 1$  gives

$$80c^2 - 200c + 9 = 80c(c - 4) + 120c + 9 < 0,$$

which is not possible, because  $c \geq 4$ ,

and  $b = 2$  gives

$$80c^2 - 288c + 144 = 80c(c - 4) + 32c + 144 < 0,$$

which is again not possible.

(iii)  $a \geq 8$ .

By the Theorem of Lagrange, since  $G$  is reduced,

$$ac \leq 4/3 d(G) = 4/3(ac - b^2),$$

so that

$$(ac - b^2) \geq 3/4ac,$$

and

$$ac(a + c - 2b) > 9/16(ac - b^2)^2$$

implies



and

$$(a + c) > \frac{81}{256}ac,$$

so that

$$\frac{1}{a} + \frac{1}{c} \geq \frac{81}{256}.$$

But

$$a \geq 8, c \geq 8, \text{ and} \\ \frac{1}{a} + \frac{1}{c} \leq \frac{1}{8} + \frac{1}{8} = \frac{1}{4} < \frac{81}{256},$$

which shows that this case is also impossible.

We have thus completed the proof of Theorem II' in case (a).

## 10. Proof of Theorem II' Case (b)

10.1 Let  $l$  be a straight line not parallel to a lattice line. Let  $\Pi$  be the plane through  $O$  perpendicular to  $l$ . Let  $\Lambda_1$  be the projection of  $\Lambda$  on  $\Pi$ . Then the lines parallel to  $l$  meet the spheres  $K + A$ ,  $A \in \Lambda$  if and only if the circles  $C + A$ ,  $A \in \Lambda_1$  cover  $\Pi$ , where  $C$  is the circle  $K \cap \Pi$ , i.e.  $C$  is the circle of radius  $3/4$ . We have then to show that every point of  $\Pi$  is within the distance  $3/4$  from some point of  $\Lambda_1$ .

If  $\text{Proj } A =$  projection of the point  $A$  of  $R^3$  on  $\Pi$ , then  $\text{Proj } (A - B) = \text{Proj } A - \text{Proj } B$ , and it follows that  $\Lambda_1$  is an additive subgroup of the group  $\Pi$  under addition. Also, since  $\Lambda$  is "three-dimensional",  $\Lambda_1$  is "two-dimensional". One can easily see that for  $\Lambda_1$ , we have the following possibilities:

- (i) If  $O$  is not a limit point of  $\Lambda_1$ , then  $\Lambda_1$  is a two-dimensional lattice, and since  $\text{Proj } (mA + nB) = m \text{Proj } A + n \text{Proj } B$ , one can easily see that  $l$  is parallel to a lattice line  $OA$  of  $\Lambda$ , and this case does not arise,
- (ii) If  $O$  is a limit point of  $\Lambda_1$ , and all points of  $\Lambda_1$  near enough to  $O$  lie on a straight line  $\alpha$  through  $O$ , then  $\Lambda_1$  is dense on  $\alpha$ , and consists of points lying dense on lines parallel to  $\alpha$  at the same distance  $\delta$  say, between consecutive ones, and
- (iii)  $\Lambda_1$  is dense everywhere in  $\Pi$ , in which case there is nothing to prove.

We have, therefore, to consider case (ii) only. In this case  $\Lambda$  is distributed in the planes orthogonal to  $\Pi$  through the lines parallel to  $\alpha$  of  $\Lambda_1$ . These planes are at a distance  $\delta$  apart (i.e. consecutive planes are at a distance  $\delta$  from each other). The part of  $\Lambda$  in the plane through  $\alpha$  is a two dimensional lattice  $\Lambda_2$  and the parts in other planes are its translates. The determinant  $d(\Lambda) = \delta \cdot d(\Lambda_2)$ , where  $d(\Lambda_2)$  is the determinant of  $\Lambda_2$ .

We notice that the squares of the distances between lattice points of  $\Lambda$  are the values of  $f = 2\sum x_i^2 - 2\sum x_i x_j$ , so that these squared distances are at least 2, and  $\Lambda$  provides a packing for spheres of radius  $(1/2)\sqrt{2}$ . Therefore,  $\Lambda_2$  provides a packing for circles of radius  $1/\sqrt{2}$ . Since the density of the closest lattice packings of circles is  $\pi/2\sqrt{3}$ , we get

$$\pi/2d(\Lambda_2) \leq \pi/2\sqrt{3}$$

and

$$d(\Lambda_2) \geq \sqrt{3}.$$

Since

$$d(\Lambda) = 2, \delta \leq 2/\sqrt{3} < 3/2.$$

Thus the distance  $\delta$  between consecutive lines parallel to  $l$  on which  $\Lambda_1$  is dense is  $< 3/2$ . Let  $P \in \Pi$ , then  $P$  is at a distance  $\leq \delta/2 < 3/4$  from one of these lines and at a distance  $< 3/4$  from some point of  $\Lambda_1$ , which completes the proof.

### Acknowledgements

This work was carried out during the visit of the first author to Ohio State University, and we are grateful to the University for making this possible. We are also grateful to Professors V C Dumir and R J Hans Gill for their assistance.

### References

- [1] Dickson L E, *Studies in the theory of numbers*, (Chicago: University Press), (1930)
- [2] Fejes Toth G, *Period Math. Hung.*, **7** (1976) 89–90
- [3] Lekkerkerker C G and Gruber P, *Geometry of numbers* (Revised edition), (Amsterdam: North-Holland), (1987)
- [4] Makai E Jr., *Stud. Sci. Math. Hung.*, **13** (1978) 19–27

# The number of ideals in a quadratic field

M N HUXLEY and N WATT

School of Mathematics, University of Wales College of Cardiff, Senghenydd Road, Cardiff  
Wales CF2 4YH, UK

Dedicated to the memory of Professor K G Ramanathan

**Abstract.** Let  $K$  be a quadratic field, and let  $R(N)$  be the number of integer ideals in  $K$  with norm at most  $N$ . Let  $\chi$  with conductor  $k$  be the quadratic character associated with  $K$ . Then

$$|R(N) - NL(1, \chi)| \leq Bk^{50/73} N^{23/73} (\log N)^{461/146}$$

for  $N \geq Ak$ , where  $A$  and  $B$  are constants. For  $N \geq Ak^C$ ,  $C$  sufficiently large, the factor  $k^{50/73}$  may be replaced by  $(d(k))^{4/73} k^{46/73}$ .

**Keywords.** Exponential sums; character sums; ideals; quadratic field.

## 1. Introduction

Let  $K$  be a quadratic number field, and let  $r(n)$  be the number of integer ideals in  $K$  whose norm is  $n$ . Then

$$r(n) = \sum_{d|n} \chi(d),$$

where  $\chi(d)$  is a real primitive character whose conductor  $k$  is the absolute value of the discriminant of  $K$ . Let  $R(N)$  be the number of integer ideals with norm at most  $N$ . Dirichlet (see [1, 6]) showed that

$$R(N) = \sum_{n \leq N} r(n) = NL(1, \chi) + O(kN^{1/2}); \quad (1.1)$$

the factor  $k$  in the error term can be reduced to  $k^{1/2} \log k$  using the Polya–Vinogradov theorem. For the Gaussian field  $Q(i)$ , the sum  $R(N)$  is the number of lattice points in a quarter-circle. The remainder term in (1.1) has been studied in this case. Recently Iwaniec and Mozzochi [5] used a powerful new method to reduce the exponent of  $N$  in (1.1) to  $7/22 + \varepsilon$ . This work was generalized and taken further by Huxley [2, 3], who obtained the error term

$$O(N^{23/73} (\log N)^{315/146}). \quad (1.2)$$

For a general quadratic field  $K$  we can apply the lattice-point method separately to each ideal class, with a remainder that depends on  $N$  as in (1.2) in the complex case, with the power of  $\log N$  increased by one in the real case, and also on the ideal class

by way of the maximum radius of curvature of the ellipse or hyperbolic segment that contains the lattice points, and in the real case, also on the fundamental unit. In this paper we obtain results that depend on  $K$  only as a power of  $k$ , even for real fields; in fact  $\chi$  can be any primitive character.

**Theorem 1.** *There are absolute constants  $A$  and  $B$  such that for  $N \geq Ak$  we have*

$$|R(N) - NL(1, \chi)| \leq Bk^{50/73} N^{23/73} (\log N)^{461/146}.$$

**Theorem 2.** *There are absolute constants  $A$ ,  $B$  and  $C$  such that for  $N \geq Ak^C$  we have*

$$|R(N) - NL(1, \chi)| \leq B(d(k))^{4/73} k^{46/73} N^{23/73} (\log N)^{461/146}.$$

The constants  $A$ ,  $B$  and  $C$  could be calculated effectively. Theorem 2 comes from an upper bound with several terms involving different powers of  $d(k)$ ,  $k$ ,  $N$  and  $\log N$ . The other terms involve smaller powers of  $N$ , but powers of  $k$  which may be closer to one. To find the infimum of those  $C$  for which Theorem 2 can be proved with some  $A$  and  $B$  would involve a large number of cases and alternative arguments.

## 2. Preparation

We obtain Theorem 1 from the following lemmas.

*Lemma 1. (Accelerated convergence for  $L(1, \chi)$ ). For any non-trivial character mod  $k$  we have*

$$L(1, \chi) = \sum_1^N \frac{\chi(n)}{n} \left(1 - \frac{n^2}{N^2}\right) + O\left(\frac{k^{3/2}}{N^2}\right).$$

*If  $\chi(-1) = 1$  and  $k|N$ , then*

$$L(1, \chi) = \sum_1^N \frac{\chi(n)}{n} + O\left(\frac{k^{3/2}}{N^2}\right).$$

*Proof.* Let

$$\rho(t) = [t] - t + \frac{1}{2}, \quad \sigma(t) = \int_0^t \rho(x) dx.$$

Since  $\sigma(t)$  has a Lipschitz condition, the weighted Polya-Vinogradov bound gives

$$\sum_{a \bmod k} \chi(a) \sigma\left(\frac{x-a}{k}\right) = O(\sqrt{k})$$

uniformly in  $x$ . Now

$$\sum_{N+1}^M \frac{\chi(n)}{n} = \sum_{a \bmod k} \chi(a) \int_{N+1/2}^{M+1/2} \frac{1}{x} d\left[\frac{x-a}{k}\right]$$

$$\begin{aligned}
 &= \sum_a \chi(a) \int_{N+1/2}^{M+1/2} \left( \frac{1}{kx} dx + \frac{1}{x} d\rho \left( \frac{x-a}{k} \right) \right) \\
 &= \sum_a \chi(a) \left[ \frac{1}{x} \rho \left( \frac{x-a}{k} \right) \right]_{N+1/2}^{M+1/2} + \sum_a \chi(a) \int_{N+1/2}^{M+1/2} \frac{1}{x^2} \rho \left( \frac{x-a}{k} \right) dx \\
 &= -\frac{1}{N+1/2} \sum_a \chi(a) \rho \left( \frac{N+1/2-a}{k} \right) + O \left( \frac{\sqrt{k \log k}}{M} \right) \\
 &\quad + \left[ \frac{k}{x^2} \sum_a \chi(a) \sigma \left( \frac{x-a}{k} \right) \right]_{N+1/2}^{M+1/2} + \int_{N+1/2}^{M+1/2} \frac{2k}{x^3} \sum_a \chi(a) \sigma \left( \frac{x-a}{k} \right) dx.
 \end{aligned} \tag{2.1}$$

For large  $M$  all terms after the first term on the right of (2.1) are  $O(k^{3/2}/N^2)$ . Similarly

$$\begin{aligned}
 \sum_1^N \frac{n\chi(n)}{N^2} &= \sum_{a \bmod k} \chi(a) \int_0^{N+1/2} \frac{x}{N^2} d \left[ \frac{x-a}{k} \right] \\
 &= \sum_a \chi(a) \frac{1}{N^2} \int_0^{N+1/2} \left( \frac{x dx}{k} + x d\rho \left( \frac{x-a}{k} \right) \right) \\
 &= \frac{1}{N^2} \sum_a \chi(a) \left[ x\rho \left( \frac{x-a}{k} \right) - k\sigma \left( \frac{x-a}{k} \right) \right]_0^{N+1/2} \\
 &= \frac{N+1/2}{N^2} \sum_a \chi(a) \rho \left( \frac{N+1/2-a}{k} \right) + O \left( \frac{k^{3/2}}{N^2} \right),
 \end{aligned}$$

which cancels with (2.1) up to  $O(k^{3/2}/N^2)$ . If  $\chi(-1) = 1$  and  $k|N$ , then

$$\sum_1^N n\chi(n) = \sum_1^N (N-n)\chi(N-n) = \sum_1^N (N-n)\chi(n);$$

but the right hand and left hand sides sum to zero. □

**Lemma 2.** (Dissection of the remainder term). Let  $\chi(n)$  be a nontrivial character mod  $k$ , and let

$$r(n) = \sum_{d|n} \chi(d).$$

The sum function  $R(N)$  of  $r(n)$  satisfies

$$R(N) = \sum_1^N r(n) = NL(1, \chi) + R_1 + R_2 + O(\sqrt{k \log k}),$$

with

$$\begin{aligned}
 R_1 &= \sum_{d \leq \sqrt{kN}} \chi(d) \rho \left( \frac{N}{d} \right), \\
 R_2 &= \sum_{a \bmod k} \chi(a) \sum_{e \leq \sqrt{N/k}} \rho \left( \frac{N/e-a}{k} \right).
 \end{aligned}$$

*Proof.*  $R(N)$  is the sum of  $\chi(d)$  over pairs of integers  $d, e$  with  $de \leq N$ . If  $\chi(d) \neq 0$ , then  $d$  is not a multiple of  $k$ , and either  $d < ke$  or  $d > ke$ . Hence

$$\begin{aligned} R(N) &= \sum_{d \leq \sqrt{kN}} \chi(d) \sum_{d/k < e \leq N/d} 1 + \sum_{e \leq \sqrt{N/k}} \sum_{ke < d \leq N/e} \chi(d) \\ &= \sum_{d \leq \sqrt{kN}} \chi(d) \left( \frac{N}{d} - \frac{d}{k} + \rho\left(\frac{N}{d}\right) - \rho\left(\frac{d}{k}\right) \right) \\ &\quad + \sum_{e \leq \sqrt{N/k}} \sum_a \chi(a) \left( \frac{N}{ek} - \frac{a}{k} - \left(e - \frac{a}{k}\right) + \rho\left(\frac{N/e - a}{k}\right) - \rho\left(e - \frac{a}{k}\right) \right). \end{aligned}$$

Since  $\rho(e - a/k) = -\rho(a/k)$  when  $\chi(a)$  is nonzero, we have

$$\begin{aligned} R(N) &= N \sum_{d \leq \sqrt{kN}} \frac{\chi(d)}{d} \left( 1 - \frac{d^2}{kN} \right) + R_1 + R_2 \\ &\quad + \sum_{a \bmod k} \chi(a) \rho\left(\frac{a}{k}\right) \left( \sum_{e \leq \sqrt{N/k}} 1 - \sum_{\substack{d \leq \sqrt{kN} \\ d \equiv a \pmod{k}}} 1 \right). \end{aligned}$$

The first term is  $NL(1, \chi) + O(\sqrt{k})$  by Lemma 1, and the last term is  $O(\sqrt{k} \log k)$  by the Polya-Vinogradov theorem.  $\square$

There are three ways of proceeding.

1. Split  $R_1$  into sums with a condition  $n \equiv a \pmod{k}$  for  $a = 1, \dots, k-1$ , and consider each value of  $a$  separately in  $R_1$  and  $R_2$ . Theorem 1 follows at once from Theorem 4 of [3], which is Theorem 5.2.4 of [4].
2. Split  $R_1$  and  $R_2$  into  $k-1$  sums corresponding to nonzero residue classes mod  $k$  as above. They form a congruence family in the sense of Lemma 4.3.6 of [4]. A slightly better bound holds on average for the sums of a congruence family.
3. Modify the method of [3] to take the character in  $R_1$  and  $R_2$  through all the Poisson summation steps.

Method 3 should be the most powerful. However the calculations produce characters of shifted arguments. In order to separate the variables in readiness for the large sieve, we must either subdivide or endure extra Gauss sums as factors in the upper bound. Also, the rank of the bilinear form in the large sieve is multiplied by a power of  $k$ .

For methods 2 and 3 we expand  $\rho(t)$  as a finite Fourier series [4, Lemma 2.1.9]:

$$\rho(t) = \sum_{h \neq 0} \frac{c(h)}{c(0)} \frac{e(ht)}{2\pi i h} + O\left(\frac{1}{H} + \min\left(1, \frac{1}{H^3 \|t\|^3}\right)\right), \quad (2.2)$$

where  $c(h)$  are the coefficients of the second Fejer kernel, expressed in terms of binomial coefficients by:

Thus  $R_1$  can be written as

$$R_1 = \sum_{m \leq \sqrt{(kN)}} \chi(m) \sum_{h \neq 0} \frac{c(h)}{c(0)} \frac{e(hN/m)}{2\pi i h} + O\left(\frac{\sqrt{(kN)}}{H}\right) + O\left(\sum_{m \leq \sqrt{(kN)}} \min\left(1, \frac{1}{H^3 \|N/m\|^3}\right)\right). \quad (2.3)$$

Similarly

$$R_2 = \sum_{b \bmod k} \chi(b) \sum_{n \leq \sqrt{(N/k)}} \sum_{h \neq 0} \frac{c(h)}{c(0)} \frac{1}{2\pi i h} e\left(\frac{hN}{kn} - \frac{bh}{k}\right) + O\left(\frac{\sqrt{(kN)}}{H}\right) + O\left(\sum_b \sum_{n \leq \sqrt{(N/k)}} \min\left(1, \frac{1}{H^3 \|N/kn - b/k\|^3}\right)\right). \quad (2.4)$$

In (2.3) the number of values of  $m$  in a range  $M \leq m < 2M$  for which  $\|N/m\| \leq \delta$  is

$$O(\delta M + N^{23/73} (\log N)^{315/146})$$

by Theorem 2 of [3], modified in the same way that Theorem 3 of [3] was modified to produce Theorem 4. When we sum  $M$  through powers of two, then the third error term in (2.3) is

$$O\left(\frac{\sqrt{(kN)}}{H} + N^{23/73} (\log N)^{461/146}\right). \quad (2.5)$$

We obtain the same estimate for the error term in (2.4).

This account is over-simplified. It is better to divide  $R_1$  and  $R_2$  into blocks in which  $m$  or  $n$  have fixed order of magnitude,  $M \leq n < 2M$  for some  $M$ , and then to choose  $H = H(M)$  in (2.2) to be constant within the block, but different for different blocks. In each block, the error term in (2.2) can also be estimated by double exponential sums (without characters) as in [3].

### 3. Congruence families of sums

A sum  $\sum \rho(g(m))$  over some range  $M \leq m < M_2$  corresponds to the remainder term in counting lattice points  $(m, n)$  in a region partly bounded by a curve  $y = g(x)$ . Counting points with  $m \equiv \ell \pmod{k}$  corresponds to a sum

$$\sum \rho(g(km + \ell)) = \sum \rho\left(f\left(m + \frac{\ell}{k}\right)\right) \quad (3.1)$$

between suitable limits, with  $f(x) = g(kx)$ . Counting points with  $n \equiv \ell \pmod{k}$  corresponds to a sum

$$\sum \rho\left(\frac{g(m) - \ell}{k}\right) = \sum \rho\left(f(m) - \frac{\ell}{k}\right), \quad (3.2)$$

where  $f(x) = g(x)/k$ . The family of sums of the form (3.1) or (3.2) as  $\ell$  varies is called a congruence family. Theorem 8 of [3], which deals with a family of sums given by

different values of a parameter, does not apply, as the parameter must occur non-trivially, not as a linear shift. The saving occurs because changing the parameter changes the first derivative of the argument of  $\rho(t)$ . In a congruence family the change in the derivative is negligible, but the function itself changes in a predictable way. This idea was developed by Watt [7] for simple exponential sums. We obtain results that correspond to Theorems 7 and 8 of [3].

**Lemma 3.** (Congruence families of double exponential sums). Let  $F(x)$  be a real function with four continuous derivatives for  $1 \leq x \leq 2$ , and let  $g(x)$ ,  $G(x)$  be bounded functions of bounded variation on  $1 \leq x \leq 2$ . Let  $C_0, \dots, C_5$  be real numbers  $\geq 1$ . Let  $H$  and  $M$  (integers) and  $T$  (real) be large parameters. Suppose that

$$|F^{(r)}(x)| \leq C_r$$

for  $r = 1, \dots, 4$ , that

$$|F^{(r)}(x)| \geq 1/C_r \quad (3.3)$$

for  $r = 1, 2$ , and that either case 1 or case 2 holds:

Case 1.  $M \leq C_0 T^{1/2}$  and (3.3) holds for  $r = 3$  also.

Case 2.  $M \geq C_0^{-1} T^{1/2}$  and

$$|F'F^{(3)} - 3F''^2| \geq 1/C_5.$$

Let  $k$  be a fixed positive integer, and for  $\ell = 0, \dots, k-1$  let  $S_\ell$  denote either the sum

$$S_\ell = \sum_{h=H}^{2H-1} g\left(\frac{h}{H}\right) \sum_{m=M}^{2M-1} G\left(\frac{m}{M}\right) e\left(\frac{hT}{M} F\left(\frac{m}{M} + \frac{\ell}{kM}\right)\right) \quad (3.4)$$

or the sum

$$S_\ell = \sum_{h=H}^{2H-1} g\left(\frac{h}{H}\right) \sum_{m=M}^{2M-1} G\left(\frac{m}{M}\right) e\left(\frac{hT}{M} F\left(\frac{m}{M}\right) - \frac{h\ell}{k}\right). \quad (3.5)$$

Then there are constants  $C_6$ ,  $C_7$  and  $C_8$  constructed from  $C_0, \dots, C_5$  such that if

$$C_6 T^{1/3} \leq M \leq C_6^{-1} T^{2/3}$$

and

$$H \leq C_7 \min(M^{3/2}/T^{1/2}, M^{1/2}, MT^{-7/27}),$$

then we have bounds of the form

$$\sum_{\ell=0}^{k-1} |S_\ell|^2 = O(EkH^2 T(\log T)^{9/2}), \quad (3.6)$$

where the constant in the upper bound is constructed from  $C_0, \dots, C_5$ , from the bounds for the functions  $g(x)$  and  $G(x)$ .



Case (a). In cases 1 and 2, for

$$\begin{aligned} & \left( \frac{d^3(k)HT}{k^3 M} \right)^{1/5} + \left( \frac{MT^{1/8}}{H} \right)^{1/4} + \frac{d(k)}{k} \left( \frac{HT^{1/3}}{M} \right)^4 + \left( \frac{HT^{1/3}}{M} \right)^{5/2} \\ & \geq C_8^{-1} \min \left( \frac{T}{M^2} + \frac{M^2}{T}, \frac{H^5 T^2}{M^5} \right), \end{aligned} \quad (3.7)$$

we have (3.6) with

$$\begin{aligned} E &= \left( \frac{d(k)}{k} \right)^{4/35} \left( \frac{H}{M} \right)^{3/35} \frac{1}{T^{12/35}} + \frac{1}{T^{3/8}} \\ &+ \left( \frac{d(k)}{k} \right)^{1/2} \left( \frac{H}{M} \right)^{15/4} T^{3/4} + \left( \frac{H}{M} \right)^3 T^{1/2}. \end{aligned} \quad (3.8)$$

Case (b). In case 1 we have (3.6) with

$$\begin{aligned} E &= \left( \frac{d(k)}{k} \right)^{2/7} \frac{H^{1/7} M^{3/7}}{T^{4/7}} + \frac{M^{9/11}}{H^{1/11} T^{8/11}} + \frac{1}{H^{1/9} M^{1/3} T^{2/9}} \\ &+ \left( \frac{d(k)}{k} \right)^{1/2} \frac{H}{T^{1/2}} + \frac{H^{1/4} M^{3/4}}{T^{3/4}}. \end{aligned} \quad (3.9)$$

Case (c). In case 2 we have (3.6) with

$$\begin{aligned} E &= \left( \frac{d(k)}{k} \right)^{2/7} \frac{H^{1/7}}{M^{5/7}} + \frac{1}{H^{1/11} M^{7/11}} + \frac{M^{5/9}}{H^{1/9} T^{2/3}} \\ &+ \left( \frac{d(k)}{k} \right)^{1/2} \frac{HT^{1/2}}{M^2} + \frac{H^{1/4} T^{1/4}}{M^{5/4}}. \end{aligned} \quad (3.10)$$

*Proof.* The proof is a variation on that of Theorem 7 in [3]. The sum over  $m$  is divided into short intervals labelled by rational numbers (Farey arcs), with an approximate equivalence relation that we call resonance. Approximate or fuzzy equivalence means that transitivity weakens the approximation. In [3] the extra structure given by the parameter is used only to compare corresponding Farey arcs in different sums of the family. Here we use the congruence structure in the same way. The congruence structure is simpler, so there is less constraint on the length of the short intervals. The comparison occurs differently, and the possible saving is less. The second term and the last term in the bounds for  $E$  dominate in cases when the maximum saving occurs. The other terms correspond to terms and cases in Theorem 7 of [3].  $\square$

We make some changes of notation in order to apply Lemma 3 to the sums  $R_1$  and  $R_2$ . In  $R_1$  we must classify the values of  $m$  into residue classes  $\ell \pmod{k}$ , so that  $m = kn + \ell$  for some  $n$ . In  $R_2$  we merely write  $\ell$  for  $b$ . The variable  $n$  in  $R_1$  and  $R_2$  becomes  $m$  for Lemma 3. We write  $H_1$  for  $H(M)$  in (2.2), so that we can use  $H$  as a

$$F(x) = 1/x, \quad T = N/k,$$

and with  $g(x)$  a continuous function with

$$g\left(\frac{h}{H}\right) = \frac{H}{2\pi h} \frac{c(h)}{c(0)}$$

(the factor  $H$  is inserted for homogeneity), and

$$G\left(\frac{x}{M}\right) = \begin{cases} 1 & \text{for } x \leq \sqrt{(N/k)}, \\ 0 & \text{for } x > \sqrt{(N/k)}. \end{cases}$$

To prove Theorem 2 we need

$$H_1 = C_9 \left(\frac{k}{d(k)}\right)^{4/73} MT^{-23/73} (\log T)^{-315/146} \quad (3.11)$$

for some  $C_9$  (which affects the constant  $B$ ), to overcome the term  $1/H_1$  in (2.2), and

$$\sum_{\ell} |S_{\ell}|^2 = O\left(\left(\frac{d(k)}{k}\right)^{8/73} kH^2 T^{46/73} (\log T)^{315/73}\right) \quad (3.12)$$

for each block sum. There are  $O(\log^2 T)$  different block sums, for different size ranges of  $H$  and  $M$ . The various cases of Lemma 3 give ranges  $H_2(M) \leq H \leq H_3(M)$  in which (3.12) holds, actually with an extra factor of the form

$$(H_2(M)/H)^{\delta_1} + (H/H_3(M))^{\delta_2}$$

for some positive  $\delta_1$  and  $\delta_2$ . The sum over blocks of  $h$  gives a constant factor, not a logarithmic one. We always have  $M = O(\sqrt{T})$ . The terms in case (a) of Lemma 3 have  $H$  to a power greater than two. If the first term in (3.8) gives the order of magnitude of  $E$ , then (3.12) holds for  $H \leq H_1$ . The second term is smaller for

$$k/d(k) \leq C_{10} T^{3/64} (\log T)^{-9/16}, \quad (3.13)$$

and the third and fourth terms in (3.8) do not matter for  $H \leq H_1$ . For

$$C_{11} \left(\frac{k}{d(k)}\right)^{1/6} T^{7/16} \leq M \leq C_{11}^{-1} \left(\frac{d(k)}{k}\right)^{1/6} T^{9/16} \quad (3.14)$$

the condition (3.7) is satisfied for all  $H$ . For smaller  $M$  we use case (b) for small  $H$ . The order of magnitude of  $E$  changes smoothly as we pass from case (a) to case (b). The terms in (3.9) with  $H$  in the denominator may make (3.12) fail for small  $H$ . We must also consider  $H$  and  $M$  below the ranges permitted in Lemma 3.

As in [3], for small  $H$  or  $M$  we use the simple exponential sum bound from the exponent-pair  $(2/7, 4/7)$  to get

$$S_{\ell} = O(HT)^{2/7},$$

which implies (3.12) for

$$H \leq H_0 = \left( \frac{d(k)}{k} \right)^{14/73} T^{15/146} (\log T)^{2205/292}.$$

This range contains  $H \leq H_1$  for

$$M \leq C_{12} \left( \frac{d(k)}{k} \right)^{18/73} T^{61/146} (\log T)^{2835/292}. \quad (3.15)$$

We find that blocks with  $H > H_0$  satisfy (3.12) by case (b) of Lemma 3 for

$$\begin{aligned} C_{13} \left( \frac{k}{d(k)} \right)^{86/219} T^{179/438} (\log T)^{-573/292} &\leq M \\ &\leq C_{13} \left( \frac{d(k)}{k} \right)^{34/219} T^{589/1314} (\log T)^{179/292}. \end{aligned} \quad (3.16)$$

For

$$\frac{k}{d(k)} \leq C_{14} T^{1/70} (\log T)^{639/35}, \quad (3.17)$$

the range of  $M$  in (3.16) overlaps the ranges (3.14) and (3.15), and we have covered all cases. Since  $T = N/k$ , we have proved Theorem 2 for any  $C > 71$ , with  $A$  chosen so that (3.13) holds. The lower bound for  $C$  can be reduced by using deeper bounds for simple exponential sums, which would increase  $H_0$ , and relax (3.15), (3.16) and (3.17). However we must have  $C > 67/3$  to satisfy (3.13).

## References

- [1] Davenport H, Multiplicative number theory (1967) (Chicago: Markham)
- [2] Huxley M N, Exponential sums and lattice points, *Proc. London Math. Soc.* **60** (1990) 471–502  
Huxley M N, Corrigendum, *Proc. London Math. Soc.* **66** (1993) 70
- [3] Huxley M N, Exponential sums and lattice points II, *Proc. London Math. Soc.* **66** (1993) 279–301
- [4] Huxley M N, Area, lattice points and exponential sums, (to appear)
- [5] Iwaniec H and Mozzochi C J, On the circle and divisor problems, *J. Number Theory* **29** (1988) 60–93
- [6] Landau E, Vorlesungen über Zahlentheorie (1927) (Leipzig: Hirsel)
- [7] Watt N, A hybrid bound for Dirichlet  $L$ -functions on the critical line, *Proc. Amalfi Conference in analytic number theory* (Salerno U P) (1992) 387–392



# On the zeros of a class of generalised Dirichlet series—XIV

R BALASUBRAMANIAN and K RAMACHANDRA\*

Institute of Mathematical Sciences, Tharamani, Madras 600 113, India

\*School of Mathematics, Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400 005, India

Dedicated to the memory of Professor K G Ramanathan

**Abstract.** We prove a general theorem on the zeros of a class of generalised Dirichlet series. We quote the following results as samples.

**Theorem A.** Let  $0 < \theta < \frac{1}{2}$  and let  $\{a_n\}$  be a sequence of complex numbers satisfying the inequality  $\left| \sum_{m=1}^N a_m - N \right| \leq (\frac{1}{2} - \theta)^{-1}$  for  $N = 1, 2, 3, \dots$ , also for  $n = 1, 2, 3, \dots$  let  $\alpha_n$  be real and  $|\alpha_n| \leq C(\theta)$  where  $C(\theta) > 0$  is a certain (small) constant depending only on  $\theta$ . Then the number of zeros of the function

$$\sum_{n=1}^N a_n (n + \alpha_n)^{-s} = \zeta(s) + \sum_{n=1}^{\infty} (a_n (n + \alpha_n)^{-s} - n^{-s})$$

in the rectangle  $(\frac{1}{2} - \delta \leq \sigma \leq \frac{1}{2} + \delta, T \leq t \leq 2T)$  (where  $0 < \delta < \frac{1}{2}$ ) is  $\geq C(\theta, \delta) T \log T$  where  $C(\theta, \delta)$  is a positive constant independent of  $T$  provided  $T \geq T_0(\theta, \delta)$  a large positive constant.

**Theorem B.** In the above theorem we can relax the condition on  $a_n$  to  $\left| \sum_{m=1}^N a_m - N \right| \leq (\frac{1}{2} - \theta)^{-1} N^\theta$  and  $|a_N| \leq (\frac{1}{2} - \theta)^{-1}$ . Then the lower bound for the number of zeros in  $(\sigma \geq \frac{1}{2} - \delta, T \leq t \leq 2T)$  is  $> C(\theta, \delta) T \log T (\log \log T)^{-1}$ . The upper bound for the number of zeros in  $(\sigma \geq \frac{1}{2} + \delta, T \leq t \leq 2T)$  is  $O(T)$  provided  $\sum_{n \leq x} a_n = x + O_\varepsilon(x^\varepsilon)$  for every  $\varepsilon > 0$ .

**Keywords.** Generalised Dirichlet series; distribution of zeros; neighbourhood of the critical line.

## 1. Introduction

This paper ought to have been paper XII of this series. But elsewhere [5] the second author has referred to this paper as paper XIV, because there are two new additions to this series namely, On the zeros of  $\zeta'(s) - a$ , (on the zeros of a class of generalised Dirichlet series-XII) and On the zeros of  $\zeta(s) - a$ , (on the zeros of a class of generalized Dirichlet series-XIII) both of which will appear in *Acta Arithmetica* with the short titles only. The addition elsewhere of the title in the brackets have been made only for some technical convenience. In the present paper we continue the investigations of the papers III [1], IV [2], V [4], and VI [3]. Just as VI [3], was in the nature of an addendum to the earlier papers, this note is a modest progress beyond the paper

VI [3], and the previous papers. Apart from an innovation, the main change consists in replacing the old kernel  $\exp(w^{4k+2})$  by the function  $R(w) = \exp((\sin w/100)^2)$ . Thus in place of  $\Delta(\chi)$ , our new function  $\Delta_1(\chi)$  will be defined for all  $\chi > 0$  by

$$\Delta_1(\chi) = \frac{1}{2\pi i} \int_{2-i\infty}^{2+i\infty} \chi^w R(w) \frac{dw}{w}.$$

Also by moving the line of integration from  $\text{Re } w = 2$  to  $\text{Re } w = -2$ , we see that  $\Delta_1(\chi) = O(\chi^2)$  and also  $\Delta_1(\chi) = 1 + O(\chi^{-2})$  where the  $O$ -constants are absolute. As a special case of a more general theorem we prove the following two theorems.

**Theorem 1.** Let  $0 < \theta < \frac{1}{2}$  and let  $\{a_n\}$  be a sequence of complex numbers satisfying the inequality

$$\left| \sum_{m=1}^N a_m - N \right| \leq \left( \frac{1}{2} - \theta \right)^{-1}$$

for  $N = 1, 2, 3, \dots$ . Also for  $n = 1, 2, 3, \dots$  let  $\alpha_n$  be real with  $|\alpha_n| \leq C(\theta)$  where  $C(\theta) > 0$  is a certain (small) constant depending only on  $\theta$ . Then the number of zeros of the function

$$\sum_{n=1}^{\infty} a_n (n + \alpha_n)^{-s} = \zeta(s) + \sum_{n=1}^{\infty} (a_n (n + \alpha_n)^{-s} - n^{-s})$$

in the rectangle  $(\sigma \geq \frac{1}{2} - \delta, T \leq t \leq 2T)$  (where  $0 < \delta < \frac{1}{2}$ ) is  $\geq C(\theta, \delta) T \log T$ , where  $C(\theta, \delta)$  is a positive constant independent of  $T$  provided  $T \geq T_0(\theta, \delta)$ , a large constant.

**Theorem 2.** Let  $0 < \theta < \frac{1}{2}$  and  $\{a_n\}$  a sequence of complex numbers satisfying the inequalities

$$|a_N| \leq \left( \frac{1}{2} - \theta \right)^{-1} \text{ and } \left| \sum_{m=1}^N a_m - N \right| \leq \left( \frac{1}{2} - \theta \right)^{-1} N^\theta$$

for  $N = 1, 2, 3, \dots$ . Let  $\alpha_n$  be as before. Then the number of zeros of the function

$$\sum_{n=1}^{\infty} a_n (n + \alpha_n)^{-s} = \zeta(s) + \sum_{n=1}^{\infty} (a_n (n + \alpha_n)^{-s} - n^{-s})$$

in the rectangle  $(\sigma \geq \frac{1}{2} - \delta, T \leq t \leq 2T)$  (where  $0 < \delta < \frac{1}{2} - \theta$ ) is  $> C(\theta, \delta) T \log T (\log \log T)^{-1}$ , where  $C(\theta, \delta)$  is a positive constant independent of  $T$  provided  $T \geq T_0(\theta, \delta)$ , a large positive constant.

**Remark 1.** In Theorem 1 the number of zeros of the function in question in  $(\sigma \geq \frac{1}{2} + \delta,$

$T \leq t \leq 2T)$  is  $O(T)$ . But in Theorem 2 to prove a similar result, we require  $\sum_{n \leq x} a_n = x + O_\varepsilon(x^\varepsilon)$  for every  $\varepsilon > 0$ . To prove these we have to prove that the mean square of the absolute value of the function in question in  $(T, 2T)$  is  $O_\varepsilon(T^\varepsilon)$  (for every  $\varepsilon > 0$ ) on the line  $\sigma = \frac{1}{2}$ . We have then to use an idea of J E Littlewood (see Theorem 9.15 (A) on page 230 of [7]).

**Remark 2.** Let  $\{x_n\}$  and  $\{y_n\}$  be any two sequences of complex numbers and let  $0 < \lambda_1 < \lambda_2 < \dots$  and further let  $\lambda_{n+1} - \lambda_n (n = 1, 2, \dots)$  lie between two positive constants. Then

$$\frac{1}{T} \int_0^T \left( \sum_{n=1}^{\infty} x_n \lambda_n^{it} \right) \left( \sum_{n=1}^{\infty} \bar{y}_n \lambda_n^{-it} \right) dt = \sum_{n=1}^{\infty} x_n \bar{y}_n + O \left( T^{-1} \left( \sum_{n=1}^{\infty} n |x_n|^2 \right)^{1/2} \right. \\ \left. \times \left( \sum_{n=1}^{\infty} n |y_n|^2 \right)^{1/2} \right)$$

where the  $O$ -constants depend only on the constants appearing in the conditions for the sequence  $\{\lambda_n\}$ . Also  $\bar{y}_n$  denotes the complex conjugate of  $y_n$ . This fundamental result is due to H L Montgomery and R C Vaughan (see [6] for a simple proof due to the second of us). It will be very much useful for our work.

## 2. Notation

From now on we adopt the following notation. The symbol  $\Delta_1(\chi)$  is already explained. We begin by explaining two Dirichlet series

$$\sum_{n=1}^{\infty} a_n b_n \lambda_n^{-s} \text{ and } F(s) = \sum_{n=1}^{\infty} a_n b_n \mu_n^{-s}$$

satisfying the conditions (i) to (ix) below. (We nearly borrow from VI [3]. Note the following typographical corrections. In place of  $g(x)g'(x)$  in the condition (iii) on page 247 of VI [3] there should be  $g(x)g''(x)$ . Again in VI [3] page 248 line 7 from the top,  $x$  should be  $X$  and there should be extra term  $X^{1-2\sigma}(f(X))^2$  in the bracket and in line 8 from the top " $X$ " should read  $X$  and  $\sigma$  if  $\sigma \geq 0$ . Also in V [4] on page 304 line 11 from the bottom  $F(s)$  should be  $\psi$ ). Throughout we assume  $a_n = O(1)$ .

(i)  $0 < \lambda_1 < \lambda_2 < \dots$  and  $\lambda_{n+1} - \lambda_n (n = 1, 2, 3, \dots)$  should lie between two positive constants. The sequence  $\{\lambda_n\}$  is further restricted by the condition (vii) or (viii) as the case may be.

Let  $f(x)$  and  $g(x)$  be two positive real valued functions defined in  $x \geq 0$  satisfying.

(ii)  $f(x)x^\eta$  is monotonic increasing and  $f(x)x^{-\eta}$  is monotonic decreasing for every fixed  $\eta > 0$  and all  $x \geq x_0(\eta)$ .

(iii)  $\lim_{x \rightarrow \infty} (g(x)x^{-1}) = 1$ .

(iv) For all  $x \geq 0$ ,  $0 < a \leq g'(x) \leq b$  and  $0 < a \leq (g'(x))^2 - g(x)g''(x) \leq b$  where  $a$  and  $b$  are constants.

Let  $\{a_n\}$  and  $\{b_n\}$  be two sequences of complex numbers having the following properties.

(v)  $|b_n|$  lies between  $a f(n)$  and  $b f(n)$  for all  $n$ .

(vi) For all  $X \geq 1$ ,  $\sum_{x \leq n \leq 2x} |b_{n+1} - b_n| \ll f(X)$ .

We next assume that  $\{a_n\}$  and  $\{b_n\}$  satisfy at least one of the following two conditions (vii) and (viii).

(vii) Monotonicity condition. There exists an arithmetic progression  $\mathcal{A}$  such that

$$\lim_{x \rightarrow \infty} \left( x^{-1} \sum'_{n \leq x} a_n \right) = h, \quad (h \neq 0),$$

where the accent denotes the restriction of  $n$  to  $\mathcal{A}$ . Also  $|b_n| \lambda_n^{-1/400}$  is monotonic decreasing as  $n$  varies over  $\mathcal{A}$ .

(viii) Real part condition. There exists an arithmetic progression  $\mathcal{A}$  of integers such that

$$\liminf_{x \rightarrow \infty} \left( \frac{1}{x} \sum'_{x \leq \lambda_n \leq 2x, \operatorname{Re} a_n > 0} \operatorname{Re} a_n \right) > 0$$

and

$$\lim_{x \rightarrow \infty} \left( x^{-1} \sum'_{x \leq \lambda_n \leq 2x, \operatorname{Re} a_n < 0} \operatorname{Re} a_n \right) = 0$$

where the accent denotes the restriction of  $n$  to  $\mathcal{A}$ . (We can manage with  $\operatorname{Im} a_n$ ; but this is included in the condition stated since we can change  $b_n$  in  $\varphi$  on p. 304 of V [4] to  $ib_n$  in fact to  $\pm b_n$  or  $\pm ib_n$ ).

Note that any of (vii) or (viii) implies that

$$\left| \sum'_{n \leq x} a_n |b_n|^2 \lambda_n^{-2\sigma} \right| \gg \frac{x^{1-2\sigma} (f(x))^2}{1-2\sigma},$$

for  $\sigma < \frac{1}{2}$  and  $\sigma$  close to  $\frac{1}{2}$ , where the constant implied by  $\gg$  is independent of  $\sigma$ .

(ix) Finally let  $\beta(>0)$  be a constant. We write  $\lambda_n = \beta g(n)$  for  $n$  in  $\mathcal{A}$ . Otherwise  $\lambda_n$  are arbitrary but the sequence  $\{\lambda_n\}$  is subject to the condition (i), mentioned above. Next we write  $\mu_n = \lambda_n + \alpha_n$  where  $\{\alpha_n\}$  is any sequence of real numbers subject to  $|\alpha_n| \leq C_1$ ,  $C_1$  being a positive constant which is small enough. How small should  $C_1$  be will be stated later. ( $C_1$  will be independent of the constant  $\delta$  which appears from Theorem 7 onwards).

*Remark 1.* The earlier results were proved with the condition  $\lambda_n = g(n) + u_n + v_n$  (for all  $n$ ) where  $\{u_n\}$  and  $\{v_n\}$  denoted two arbitrary monotonic bounded sequences of real numbers. Since bounded monotonic sequences of real numbers are convergent (say  $u_n + v_n \rightarrow l$  as  $n \rightarrow \infty$ ) and in place of  $g(x)$ ,  $g(x) + l$  satisfies the conditions satisfied by  $g(x)$ , the results of the present paper are more general. However we use the results of the earlier papers III [1], IV [2], V [4] and VI [3].

*Remark 2.* Our new results are Theorems 7, 8 and 9 and their Corollaries.

### 3. Some preparations

We begin by stating

**Theorem 3.** Let  $F_1(s) = \sum_{n=1}^{\infty} (a_n b_n \Delta_1(T/\lambda_n) \lambda_n^{-s})$ . Then for  $0 < \sigma < \frac{1}{2}$  and  $T \geq 10$  we



where  $C_2 > 0$  is independent of  $T$ .

Also for  $1 \leq X \leq T$ , we have,

$$\frac{1}{T} \int_T^{2T} \left| \sum_{n=1}^{\infty} \left( a_n b_n \Delta_1 \left( \frac{X}{\lambda_n} \right) \lambda_n^{-\sigma - it} \right) \right|^2 dt < C_3 X^{1-2\sigma} (f(X))^2,$$

where  $0 < \sigma < \frac{1}{2}$  and  $C_3 > 0$  is independent of  $T$  and  $X$ .

We make two remarks by way of proof.

**Remark 1.** The first part of the theorem is nearly explained in V [4]. The role of  $F_3(s)$  in Lemma 7 (Here  $\max_{t \text{ in } I} |F_3(s)| > 0$  should read  $\max_{t \text{ in } I} |F_3(s)| \leq D$ ) of § 2 of paper VI [3]

is played by  $F_5(s) = \sum'_{\lambda_n \leq T} b_n \lambda_n^{-s}$  where the accent denotes the restriction of the sum to the integers  $\mathscr{A}$  occurring in the condition (vii) or (viii) as the case may be. Then the function  $F_5(s)$  possesses a  $g$ th power mean with  $g = g(\sigma) > 2$  if  $\sigma < \frac{1}{2}$  in the sense

$$\frac{1}{T} \int_T^{2T} |F_5(\sigma + it)|^g dt = O((T^{(1/2)-\sigma} f(T))^g).$$

This  $g$ th power moment is easily deducible from Lemma 6 of paper IV [2] which is quoted as Theorem 4 in paper V [4]. The rest of the proof follows V [4] except that  $\exp(-(\lambda_n/T))$  is replaced by  $\Delta_1(T/\lambda_n)$ . The first part of the Theorem 3 is first proved for  $\sigma$  close to  $\frac{1}{2}$  by the above method and then extended by convexity for all  $\sigma (0 < \sigma < \frac{1}{2})$ .

**Remark 2.** Let  $\sigma > 0$ . Then by using the theorem of Montgomery and Vaughan [6] quoted already we see that the LHS of the second inequality of Theorem 3 is

$$\leq C_4 \left( \sum_{\lambda_n \leq X} (f(n))^2 n^{-2\sigma} + X^2 \sum_{\lambda_n \geq X} (f(n))^2 n^{-2\sigma-2} + \frac{X^2}{T} \sum_{\lambda_n \geq X} (f(n))^2 n^{-2\sigma-1} \right).$$

Using the fact that  $f(n)n^\eta$  is monotonic increasing and  $f(n)n^{-\eta}$  is monotonic decreasing for all fixed  $\eta > 0$  and all  $n \geq n_0(\eta)$ , we see that the theorem is proved.

Note that if  $0 < \mu < \frac{1}{2} - \sigma$  we have

$$\begin{aligned} X^{1-2\sigma} (f(X))^2 &= X^{1-2\sigma-2\mu} (f(X) X^\mu)^2 \leq X^{1-2\sigma-2\mu} (f(T) T^\mu)^2 \\ &\leq \left( \frac{X}{T} \right)^{1-2\sigma-2\mu} T^{1-2\sigma} (f(T))^2 \end{aligned}$$

and so the RHS of the second inequality of Theorem 3 is

$$O \left( \left( \frac{X}{T} \right)^{1-2\sigma-2\mu} T^{1-2\sigma} (f(T))^2 \right)$$

$D(0 < D < 1)$ , not to be confused with  $D$  occurring in Remark 1 below Theorem 3, is a small constant. In that case this expression is  $O(D^{(1/2)-\sigma}(f(T) T^{(1/2)-\sigma})^2)$ , where the  $O$ -constant depends only on  $\sigma$ . Note also that the second part of Theorem 3 uses only the properties  $0 < \lambda_1 < \lambda_2 < \dots$  and  $1 \ll \lambda_{n+1} - \lambda_n \ll 1$  of  $\{\lambda_n\}$ . We now state a Corollary to Theorem 3.

**Theorem 4.** *Let*

$$F_2(s) = \sum_{n=1}^{\infty} \left( a_n b_n \left( \Delta_1 \left( \frac{T}{\lambda_n} \right) - \Delta_1 \left( \frac{DT}{\lambda_n} \right) \right) \lambda_n^{-s} \right)$$

where  $D(0 < D < 1)$  is a sufficiently small positive constant. Then if  $0 < \sigma < \frac{1}{2}$ , we have,

$$\frac{1}{T} \int_T^{2T} |F_2(\sigma + it)| dt > C_5 T^{1/2-\sigma} f(T)$$

and

$$\frac{1}{T} \int_T^{2T} |F_2(\sigma + it)|^2 dt < C_6 T^{1-2\sigma} (f(T))^2,$$

where  $C_5(> 0)$  and  $C_6(> 0)$  are independent of  $T$ .

#### 4. Main results

We now proceed to prove the analogue of Theorem 4 where  $\{\lambda_n\}$  is replaced by  $\{\mu_n\}$ .

**Theorem 5.** *Let*

$$F_3(s) = \sum_{n=1}^{\infty} \left( a_n b_n \left( \Delta_1 \left( \frac{T}{\mu_n} \right) - \Delta_1 \left( \frac{DT}{\mu_n} \right) \right) \mu_n^{-s} \right),$$

where  $D$  is the positive constant occurring in  $F_2(s)$ . Then if  $0 < \sigma < \frac{1}{2}$ , we have,

$$\frac{1}{T} \int_T^{2T} |F_3(\sigma + it)| dt > C_7 T^{1/2-\sigma} f(T)$$

and

$$\frac{1}{T} \int_T^{2T} |F_3(\sigma + it)|^2 dt < C_8 (T^{1/2-\sigma} f(T))^2,$$

where  $C_7(> 0)$  and  $C_8(> 0)$  are independent of  $T$ , provided  $C_1(> 0)$  of condition (ix) in § 2 is sufficiently small.

**Remark.** Our proof gives this theorem where the constants depend on  $\sigma$  but uniformly in a certain range for  $\sigma$  in  $\sigma < \frac{1}{2}$ . By convexity, the theorem can be upheld for all  $\sigma < \frac{1}{2}$  uniformly in  $\sigma \leq \frac{1}{2} - C_9$  where  $C_9(> 0)$  is any constant less than  $\frac{1}{2}$ . We can even secure  $C_1$  to be independent of  $C_9$ , but  $C_7$  and  $C_8$  depend on  $C_9$ .

**Proof.** The second inequality follows by the well-known Montgomery-Vaughan

theorem (see [6]), and now Theorem 5. The first part can be deduced from that of Theorem 4 as follows. Put

$$\phi(u) = \left( \Delta_1 \left( \frac{T}{\lambda_n + u} \right) - \Delta_1 \left( \frac{DT}{\lambda_n + u} \right) \right) (\lambda_n + u)^{-s}.$$

Then

$$\begin{aligned} \Delta_1 \left( \frac{T}{\mu_n} \right) - \Delta_1 \left( \frac{DT}{\mu_n} \right) \mu_n^{-s} - \phi(0) &= \int_0^{\alpha_n} \phi'(u) du \\ &= \int_{-C_1}^{C_1} Ch(u, \alpha_n) \phi'(u) du, \end{aligned}$$

where we define  $Ch(u, \alpha_n)$  to be 1 if  $u$  lies in  $(0, \alpha_n)$  if  $\alpha_n > 0$  or  $(\alpha_n, 0)$  if  $\alpha_n < 0$ . Otherwise we define  $Ch(u, \alpha_n) = 0$ . Note that

$$\begin{aligned} \phi'(u) &= -\frac{s(\lambda_n + u)^{-s-1}}{2\pi i} \int_{2-i\infty}^{2+i\infty} \left\{ \left( \frac{T}{\lambda_n + u} \right)^w - \left( \frac{DT}{\lambda_n + u} \right)^w \right\} R(w) \frac{dw}{w} \\ &\quad - \frac{(\lambda_n + u)^{-s}}{2\pi i} \int_{2-i\infty}^{2+i\infty} \left\{ \frac{T^w}{(\lambda_n + u)^{w+1}} - \frac{(DT)^w}{(\lambda_n + u)^{w+1}} \right\} R(w) dw. \end{aligned}$$

Now

$$F_3(s) - F_2(s) = \int_{-C_1}^{C_1} \left( \sum_{n=1}^{\infty} a_n b_n Ch(u, \alpha_n) \phi'(u) \right) du$$

and so

$$\frac{1}{T} \int_T^{2T} |F_3(s) - F_2(s)| dt \leq \int_{-C_1}^{C_1} \left( \frac{1}{T} \int_T^{2T} \left| \sum_{n=1}^{\infty} (a_n b_n Ch(u, \alpha_n) \phi'(u)) \right|^2 dt \right)^{1/2} du.$$

We write  $\phi'(u) = (\phi_1(u) + \phi_2(u))(\lambda_n + u)^{-s}$  with an obvious meaning for  $\phi_1(u)$  and  $\phi_2(u)$ . We have  $\phi_1(u) = O(T/\lambda_n \min((T/\lambda_n)^2, (T/\lambda_n)^{-2}))$  and  $\phi_2(u) = O(\min(T^2/\lambda_n^3, T^{-2}/\lambda_n^{-1}))$ , by moving the line of integration to  $\operatorname{Re} w = 2$  and  $\operatorname{Re} w = -2$ . We now prove that, for  $|u| \leq C_1$  there holds uniformly in  $u$ ,

$$\frac{1}{T} \int_T^{2T} \left| \sum_{n=1}^{\infty} a_n b_n \beta_n (\lambda_n + u)^{-s} \right|^2 dt \ll T^{1-2\sigma} (f(T))^2$$

where  $\beta_n$  depends only on  $n$ ,  $T$  and  $u$  and further  $\beta_n = O(\min(T^3/\lambda_n^3, \lambda_n/T))$  and  $\beta_n = O(\min(T^2/\lambda_n^3, \lambda_n/T^2))$ . Clearly the second estimate is smaller by a factor  $O(1/T)$  and hence it suffices to ignore it. By the well-known Montgomery-Vaughan theorem (see [6]) we see that LHS is

$$O \left( \sum_{n \leq T} (f(n))^2 n^{-2\sigma} \frac{n}{T} + \sum_{n \geq T} (f(n))^2 n^{-2\sigma} \frac{T^5}{n^5} \right) = O(T^{1-2\sigma} (f(T))^2).$$

Here the  $O$ -constant is independent of  $C_1$  if  $C_1$  is chosen to be smaller than a constant  $C^*( > 0)$ . This completes the proof that

$$\frac{1}{T} \int_T^{2T} |F_3(s) - F_2(s)| dt = O(C_1 T^{(1/2)-\sigma} f(T))$$

where the  $O$ -constant is independent of  $C_1$ . This completes the proof of Theorem 5.

**Theorem 6.** There are  $\gg T$  distinct integers  $M$  in  $(T, 2T)$  for each of which there holds

$$\int_M^{M+1} |F_3(\sigma + it)| dt \gg T^{(1/2)-\sigma} f(T)$$

provided  $0 < \sigma < \frac{1}{2}$ .

*Proof.* Divide  $[T, 2T]$  into intervals  $G$  of unit length ignoring a bit at one end. Put  $\Lambda(G) = \int_G |F_3(\sigma + it)| dt$  and  $Q = T^{(1/2)-\sigma} f(T)$ . Theorem 5 gives

$$\sum_G \Lambda(G) \gg TQ \text{ and } \sum_G (\Lambda(G))^2 \ll TQ^2.$$

This leads to Theorem 6.

**Theorem 7.** Suppose that  $F(s)$  defined in  $\sigma > 1$  by

$$F(s) = \sum_{n=1}^{\infty} a_n b_n \mu_n^{-s}$$

can be continued analytically in  $(\sigma \geq \frac{1}{2} - 2\delta, T \leq t \leq 2T)$  and there  $\max |F(s)| \leq T^B$  where  $B(>0)$  is a constant. Then there are  $\gg T(\log \log T)^{-1}$  distinct integers  $M$  in  $(T, 2T)$  for each of which there holds

$$\int_M^{M+1} |F(\sigma + it)| dt \gg T^{(1/2)-\sigma} f(T)$$

where  $\sigma = \frac{1}{2} - \delta$ .

Using Theorem 3 of paper III [1] we obtain the following Corollary.

# COROLLARY

$F(s)$  has  $\gg T \log T (\log \log T)^{-1}$  zeros in  $(\sigma \geq \frac{1}{2} - 2\delta, T \leq t \leq 2T)$ .

*Remark.* It is not hard to prove that in many cases (for example  $\sum_{n \geq x} a_n = x + O_\varepsilon(x^\varepsilon)$  for every  $\varepsilon > 0$ ) that

$$\frac{1}{T} \int_T^{2T} \left| F\left(\frac{1}{2} + it\right) \right|^2 dt \ll_\varepsilon T^\varepsilon$$

(for every  $\varepsilon > 0$ ) and in this case it follows that the number of zeros of  $F(s)$  in  $(\sigma \geq \frac{1}{2} + \delta, T \leq t \leq 2T)$  is  $O(T)$ .

*Proof.* (Of Theorem 7). We have for  $s = \sigma + it$

$$F_3(s) = \frac{1}{2\pi i} \int_{2-i\infty}^{2+i\infty} F(s+w)(T^w - (DT)^w) R(w) \frac{dw}{w}.$$

We deform the contour  $(2 - i\infty, 2 + i\infty)$  to  $(2 - i\infty, 2 - iC_{10} \log \log T, -iC_{10} \log \log T,$

$iC_{10} \log \log T$ ,  $2 + iC_{10} \log \log T$ ,  $2 + i\infty$ ), use  $|F_3(s)| \leq \int |\dots| dw/w$  (over the new contour) and integrate with respect to  $t$  from  $M$  to  $M+1$  of Theorem 6. We obtain the theorem by slight work. (Here  $C_{10} (> 10)$  is a large constant).

The remark below the Corollary to Theorem 7 follows from an idea of J E Littlewood (see Theorem 9.15(A) on page 230 of [7]).

**Theorem 8.** *We have, for  $\sigma < \frac{1}{2}$ ,*

$$\frac{1}{T} \int_T^{2T} |F(\sigma + it)| dt \gg T^{(1/2) - \sigma} f(T).$$

*Proof.* Let  $T + (T/10) \leq t \leq 2T - (T/10)$  and  $\sigma < \frac{1}{2}$ . We start with the formula for  $F_3(s)$  as in the proof of Theorem 7 above and deform the contour exactly as before. It follows that

$$\frac{1}{T} \int_{T - C_{10} \log \log T + (T/10)}^{2T + C_{10} \log \log T - (T/10)} |F(\sigma + it)| dt \gg T^{(1/2) - \sigma} f(T)$$

on using the first part of Theorem 5. For this we need

$$\frac{1}{T} \int_{T + T/10}^{2T - T/10} |F_3(\sigma + it)| dt \gg T^{(1/2) - \sigma} f(T).$$

But this can be proved just as we proved the first part of Theorem 7. This completes the proof of Theorem 8.

**Theorem 9.** *Let  $\sum_{n \leq x} a_n = O(1)$ . Then for  $0 < \sigma < \frac{1}{2}$ , we have,*

$$\frac{1}{T} \int_T^{2T} |F(\sigma + it)|^2 dt \ll T^{1 - 2\sigma} (f(T))^2$$

and

$$\frac{1}{T} \int_T^{2T} \left| F\left(\frac{1}{2} + it\right) \right|^2 dt \ll_\varepsilon T^\varepsilon$$

for every  $\varepsilon > 0$ .

#### COROLLARY

Let  $0 < \sigma < \frac{1}{2}$ . Then there are  $\gg T$  distinct integers  $M$  in  $(T, 2T)$  for each of which there holds

$$\int_M^{M+1} |F(\sigma + it)| dt \gg T^{(1/2) - \sigma} f(T).$$

Hence as before  $F(s)$  has  $\gg T \log T$  zeros in  $(\sigma \geq \frac{1}{2} - \delta, T \leq t \leq 2T)$  and only  $O(T)$  zeros in  $(\sigma \geq \frac{1}{2} + \delta, T \leq t \leq 2T)$ .

We remark finally that Theorems 7, 8 and 9 are valid even if we omit  $N$  terms (other than the first term) in  $F(s)$  where  $N = O_\varepsilon(T^\varepsilon)$  for every  $\varepsilon > 0$ .

P.S. In a forthcoming paper (On the zeros of a class of generalised Dirichlet series-XV) we consider zeros of functions like  $\sum_{n=1}^{\infty} d(n)(n+\alpha_n)^{-s}$  and  $\sum_{n=1}^{\infty} d_3(n)(n+\alpha_n)^{-s}$  and prove some interesting lower bounds for the number of zeros in  $(\sigma \geq \frac{1}{2} - \delta, T \leq t \leq 2T)$  like  $\gg T \log T$ . Also in paper XVI with the same title K Ramachandra and A Sankaranarayanan have proved the upper bound  $\ll T$  in  $(\sigma \geq \frac{1}{2} + \delta, T \leq t \leq 2T)$  for the functions such as those mentioned above.

## References

- [1] Balasubramanian R and Ramachandra K, On the zeros of a class of generalised Dirichlet series-III, *J. Ind. Math. Soc.*, **41** (1977) 301-315
- [2] Balasubramanian R and Ramachandra K, On the zeros of a class of generalised Dirichlet series-IV, *J. Ind. Math. Soc.*, **42** (1978) 135-142
- [3] Balasubramanian R and Ramachandra K, On the zeros of a class of generalised Dirichlet series-VI, *Ark. Mat.*, **19** (1981) 239-250.
- [4] Ramachandra K, On the zeros of a class of generalised Dirichlet series-V, *J. Reine Angew. Math.*, **303/304** (1978) 295-313
- [5] Ramachandra K, A brief report on the zeros of a class of generalised Dirichlet series, (Proceedings of an International Conference in Kyoto held during 10-13 November 1992) ed. S Kanemitsu (1993) 48-56
- [6] Ramachandra K, Some remarks on a theorem of Montgomery and Vaughan, *J. Number Theory*, **11** (1979) 465-471
- [7] Titchmarsh E C, The theory of the Riemann zeta-function, Second edition (revised and edited by Heath-Brown D R) (Oxford: Clarendon Press) (1986)

## Local zeta functions of general quadratic polynomials\*

JUN-ICHI IGUSA

Department of Mathematics, The Johns Hopkins University, Baltimore, Maryland 21218,  
USA

Dedicated to the memory of Professor K G Ramanathan

**Abstract.** This paper is concerned with the kind of local zeta functions now often called Igusa's local zeta functions: A simple closed form of such a zeta function for an arbitrary quadratic form, its variants, and an application are given.

**Keywords.** Local zeta function; quadratic polynomials.

### 1. Introduction

The local zeta functions which we shall consider in this paper are of the form

$$Z(s) = \int_L |f(x)|_K^s dx, \quad \operatorname{Re}(s) > 0,$$

where  $L$  is a lattice in a vector space  $V$  over a  $p$ -adic field  $K$ ,  $|\cdot|_K$  is an absolute value on  $K$ ,  $f$  is a polynomial function on  $V$ ,  $s$  is a complex variable, and  $dx$  is the Haar measure on  $V$  normalized as  $Z(0) = 1$ . If  $q$  is the cardinality of the residue class field of  $K$  and if  $\operatorname{char}(K) = 0$ , our general theorem states that  $Z(s)$  is a rational function of  $q^{-s}$ ; cf. [2]. Furthermore  $Z(s)$  has been computed in a large number of cases. We might add that in some cases computations are very difficult and that no algorithm to compute  $Z(s)$  for a general  $f$  is known.

Now, in the difficult cases where  $Z(s)$  has been computed, one has often encountered auxiliary integrals of the above form where  $f$  is a quadratic polynomial with coefficients depending on some parameters. In this paper we shall give a simple closed form to  $Z(s)$  in the case where  $f$  is an arbitrary nondegenerate quadratic form and also in some inhomogeneous cases; for the sake of simplicity we have assumed that  $q$  is odd. The formula has turned out to be useful for the computation of  $Z(s)$  in some new cases and for the simplification in some known cases. We have included just one application at the end of the paper.

### 2. Preliminaries

As in the Introduction, we denote by  $V$  a (finite dimensional) vector space over a  $p$ -adic field  $K$  and by  $L$  a lattice in  $V$ . If  $S$  is any subset of  $V$ , we shall denote by  $KS$  the  $K$ -span of  $S$  in  $V$ . If  $\bar{O}_K$  is the maximal compact subring of  $K$ , then  $L$  is a free

$O_K$ -submodule of  $V$  such that  $KL = V$ . We shall denote by  $n(L)$  the rank of  $L$ , i.e.,  $n(L) = \dim(KL)$ . We choose  $\pi$  from  $O_K$  such that  $\pi O_K$  becomes the maximal ideal of  $O_K$  and denote by  $q$  the cardinality of the residue class field  $O_K/\pi O_K$ , i.e.,  $O_K/\pi O_K = \mathbb{F}_q$ . In general if  $R$  is any associative ring with 1, we shall denote by  $R^\times$  the group of units of  $R$ . We shall assume, once and for all, that  $q$  is odd; then

$$(O_K^\times)/(O_K^\times)^2 \simeq \mathbb{F}_q^\times/(\mathbb{F}_q^\times)^2 \simeq \{\pm 1\}.$$

We shall denote the corresponding homomorphism of  $O_K^\times$  to  $\{\pm 1\}$  by  $\chi$ .

We take a quadratic form  $Q$  on  $V$ , i.e., a  $K$ -valued function on  $V$  satisfying  $Q(ax) = a^2 Q(x)$  for every  $a$  in  $K$  and  $x$  in  $V$  such that  $Q(x, y) = Q(x + y) - Q(x) - Q(y)$  is  $K$ -bilinear in  $x, y$ . If  $\{w_1, \dots, w_n\}$  is a  $K$ -basis for  $V$  so that  $n = \dim(V)$ , then

$$Q\left(\sum_{1 \leq i \leq n} x_i w_i\right) = \sum_{1 \leq i \leq n} Q(w_i) x_i^2 + \sum_{i < j} Q(w_i, w_j) x_i x_j$$

for every  $x_1, \dots, x_n$  in  $K$ . Therefore if  $\{w_1, \dots, w_n\}$  is an  $O_K$ -basis for  $L$ , then  $Q$  is  $O_K$ -valued on  $L$  if and only if the coefficients of the above homogeneous polynomial of degree 2 in  $x_1, \dots, x_n$  are all in  $O_K$ . We shall assume that  $Q$  is nondegenerate, i.e., that  $Q(x, y) = 0$  for all  $y$  in  $V$  implies  $x = 0$ . If  $\{w_1, \dots, w_n\}$  is an  $O_K$ -basis for  $L$  and if  $h$  is the square matrix of degree  $n$  with  $Q(w_i, w_j)$  as its  $(i, j)$ -entry for  $1 \leq i, j \leq n$ , then  $\det(h) \neq 0$ . We define the discriminant  $D(L, Q)$  of  $(L, Q)$  as

$$D(L, Q) = (-1)^{n(n-1)/2} \cdot \det(h).$$

We observe that  $D(L, Q)(O_K^\times)^2$  is independent of the choice of the  $O_K$ -basis for  $L$ . We call  $(L, Q)$  *unimodular* if  $Q$  is  $O_K$ -valued on  $L$  and if  $D(L, Q)$  is in  $O_K^\times$ . In the general case  $(L, Q)$  has the following *Jordan decomposition*:

There exist  $O_K$ -submodules  $L_1, \dots, L_t$  of  $L$  with  $L$  as their sum and a sequence of integers  $e_1 < \dots < e_t$  such that  $KL_1, \dots, KL_t$  are mutually orthogonal with respect to  $Q(x, y)$  and  $(L_i, Q_i)$ , where  $Q_i = \pi^{-e_i} Q|_{KL_i}$ , is unimodular for  $1 \leq i \leq t$ . Furthermore if we put  $n_i = n(L_i)$ , then

$$\text{inv}(L, Q) = \{n_i, e_i, \chi(D(L_i, Q_i)); 1 \leq i \leq t\}$$

characterizes the isomorphism class of  $(L, Q)$ .

We refer to O'Meara [5], Chapter IX for the theory of Jordan decompositions. We keep in mind that  $M = L_{j_1} + \dots + L_{j_r}$ , where  $1 \leq j_1 < \dots < j_r \leq t$ , gives the Jordan decomposition of  $(M, Q|_{KM})$ .

### 3. $Z(s)$ for a quadratic form

We start from the Jordan decomposition of  $(L, Q)$  recalled in the previous section. If we define  $Q^0$  as  $Q_1 + \dots + Q_t$ , i.e., as

$$Q^0(x_1 + \dots + x_t) = Q_1(x_1) + \dots + Q_t(x_t)$$

for all  $x_1, \dots, x_t$  respectively in  $KL_1, \dots, KL_t$ , then  $(L, Q^0)$  is unimodular. We put



and  $\chi(L, Q) = 0$  otherwise. In view of the formal identity

$$\left(\sum_i n_i\right)\left(\sum_i n_i - 1\right) = \sum_i n_i(n_i - 1) + 2 \cdot \sum_{i < j} n_i n_j,$$

if  $n \equiv \sum n_i e_i \equiv 0 \pmod{2}$ , then

$$\chi(L, Q) = \chi(-1)^{\sum_{i < j} n_i n_j} \cdot \prod_{1 \leq i \leq t} \chi(D(L_i, Q_i)).$$

We observe that  $\chi(L, Q)$  is not only independent of the choice of the  $O_K$ -basis for  $L$ , but also it remains invariant under  $Q \mapsto u_1 Q$  and  $\pi \mapsto u_2 \pi$  for any  $u_1, u_2$  in  $O_K^\times$ . This follows from the fact that  $D(L, Q^0)$  is multiplied, for a fixed  $O_K$ -basis for  $L$ , by  $u^n$  under  $Q \mapsto uQ$  and by  $u^{-e}$ , where  $e = \sum n_i e_i$ , under  $\pi \mapsto u\pi$  for every  $u$  in  $O_K^\times$ . In the special case where  $e_1 \equiv \dots \equiv e_t \pmod{2}$  we shall write  $\chi(L)$  instead of  $\chi(L, Q)$ . In other words if  $e_1 \equiv \dots \equiv e_t \pmod{2}$ , then we define  $\chi(L)$  as follows:

$$\chi(L) = \begin{cases} \chi(D(L, Q^0)) & n(L) \text{ even} \\ 0 & n(L) \text{ odd.} \end{cases}$$

The notation  $\chi(M)$  will be used for

$$M = L_{j_1} + \dots + L_{j_r},$$

where  $1 \leq j_1 < \dots < j_r \leq t$  and  $e_{j_1} \equiv \dots \equiv e_{j_r} \pmod{2}$ , relative to  $Q|KM$  with the understanding that  $\chi(M) = 1$  for  $r = 0$ .

In the Introduction we have defined  $Z(s)$  as

$$Z(s) = \int_L |f(x)|_K^s dx, \quad \operatorname{Re}(s) > 0.$$

We shall normalize the absolute value  $|\cdot|_K$  on  $K$  as  $|\pi|_K = q^{-1}$  so that  $|a|_K$  for every  $a$  in  $K^\times$  becomes the module of the multiplication by  $a$  in  $K$ . Furthermore, for the sake of simplicity, we put

$$[a, b] = 1 - q^{-(a+bs)}, \quad [a, b]_+ = 1 + q^{-(a+bs)}, \quad [a] = [a, 0], \quad [a]_+ = [a, 0]_+,$$

in which  $a, b$  are mostly nonnegative integers.

**Theorem 1.** *The local zeta function  $Z(s)$  defined as*

$$Z(s) = \int_L |Q(x)|_K^s dx, \quad \operatorname{Re}(s) > 0$$

*has the following closed form in terms of  $\operatorname{inv}(L, Q)$ : For each  $i$ ,  $1 \leq i \leq t$  put*

$$Z_i = [1][0, 1]/[1, 1][m_1, 2][m_2, 2] \cdot q^{-1/2 \cdot \sum_{j < i} n_j(e_i - e_j) - e_i s} \\ \cdot \{\chi(M_1)[n_i]q^{-m_1 - s} + \chi(M_2)[m_2, 2]q^{-m_1/2} - \chi(M_3)[m_1, 2]q^{-m_2/2}\},$$

where

$$M_1 = \sum_{j < i, e_j \not\equiv e_i} L_j, \quad M_2 = \sum_{j < i, e_j \equiv e_i} L_j, \quad M_3 = M_2 + L_i,$$

$$m_1 = \sum_{j < i} n_j, \quad m_2 = m_1 + n_i.$$

Then

$$Z(s) = Z_1 + \dots + Z_t.$$

In the above definition of  $M_1, M_2$  the congruences are mod 2. We observe that the condition " $e_{j_1} \equiv \dots \equiv e_{j_r} \pmod{2}$ " is satisfied by each  $M_k$ , hence  $\chi(M_k)$  is defined in terms of  $\text{inv}(M_k, Q|KM_k)$ , hence of  $\text{inv}(L, Q)$ . In fact, if  $L_j$  occurs in  $M_k$ , then  $e_j \equiv e_i + 1, e_i, e_i \pmod{2}$  respectively for  $k = 1, 2, 3$ .

#### 4. Some lemmas

We first state a " $p$ -adic stationary phase formula," abbreviated as SPF, as a lemma:

*Lemma 1.* Let  $f(x)$  denote an element of the polynomial ring  $O_K[x_1, \dots, x_n]$  and  $\bar{f}(x)$  its image in  $F_q[x_1, \dots, x_n]$  under the canonical homomorphism  $O_K \rightarrow F_q$ ; let  $\bar{f}^{-1}(0)$  denote the subset of  $F_q^n$  defined by  $\bar{f}(\xi) = 0$ ,  $\bar{S}$  its subset defined by

$$(\partial \bar{f} / \partial x_1)(\xi) = \dots = (\partial \bar{f} / \partial x_n)(\xi) = 0,$$

and  $S$  the preimage of  $\bar{S}$  under  $O_K^n \rightarrow F_q^n$ ; put  $N = \text{card}(\bar{f}^{-1}(0))$ . Then

$$\int_{O_K^n} |f(x)|_K^s dx = (1 - q^{-n}N) + (N - \text{card}(\bar{S}))[1]q^{-n-s}/[1, 1] + \int_S |f(x)|_K^s dx.$$

We refer to [4] for the proof of a more general statement. If  $f(x)$  is homogeneous of degree  $d \geq 1$  and if  $\bar{S} = \{0\}$ , then SPF gives

$$\int_{O_K^n} |f(x)|_K^s dx = \{(1 - q^{-n}N)[1, 1] + (N - 1)[1]q^{-n-s}\}/[1, 1][n, d].$$

If further  $d = 2$ , i.e., if  $(O_K^n, Q)$  is unimodular for  $Q = f$ , then  $N$  has the following well-known expression:

$$N = \begin{cases} q^{n-1} + \chi(D)[1]q^{n/2} & n \text{ even} \\ q^{n-1} & n \text{ odd,} \end{cases}$$

in which  $D = D(O_K^n, Q)$ ; cf., e.g., Bourbaki [1]. If we use our notation that  $\chi(L)$  for  $L = O_K^n$  represents  $\chi(D)$  or 0 according as  $n$  is even or odd, then we get

$$\int_L |Q(x)|_K^s dx = [1]\{1 - \chi(L)[0, 1]q^{-n/2} - q^{-n-s}\}/[1, 1][n, 2],$$

which is also well known (in the usual notation).

For the sake of clarity we have separated the following lemmas from the proof of Theorem 1:

for  $i = 1, 2$  and  $A_{12}$  similarly as

$$A_{12} = [1] \{1 - \chi_{12}[0, 1]q^{-(n_1+n_2)/2} - q^{-n_1-n_2-s}\} / [1, 1][n_1 + n_2, 2],$$

where  $\chi_1, \chi_2, \chi_{12}$  are variables, then

$$\begin{aligned} [n_1 + n_2, 2](A_{12} - A_1) &= [1][0, 1]q^{-n_1/2} / [1, 1][n_1, 2] \\ &\cdot \{[n_2]q^{-n_1/2-s} + \chi_1[n_1 + n_2, 2] - \chi_{12}[n_1, 2]q^{-n_2/2}\}, \\ ([n_1, 2] - [n_1 + n_2, 2])A_1 + [n_2, 2]q^{-n_1-s}A_2 &= [1][0, 1]q^{-n_1-s} / [1, 1][n_1, 2] \\ &\cdot \{\chi_1[n_2]q^{-n_1/2-s} + [n_1 + n_2, 2] - \chi_2[n_1, 2]q^{-n_2/2}\}. \end{aligned}$$

**Lemma 3.** If  $a, b, c$  are integers and  $x, y$  are variables, then

$$(1 - x^b)(1 - x^{a+b+c}y^2) - (1 - x^{b+c})(1 - x^{a+b}y^2) + x^b(1 - x^c)(1 - x^ay^2) = 0,$$

hence  $[b][a + b + c, 2] - [b + c][a + b, 2] + q^{-b}[c][a, 2] = 0$ .

The verifications are all straightforward. In Lemma 2 we shall take  $\chi(L_i)$  as  $\chi_i$  for  $i = 1, 2$  and  $\chi(L_1 + L_2)$  as  $\chi_{12}$ , this in the case where  $e_1 \equiv e_2 \pmod{2}$ .

**5. Proof of Theorem 1.** If  $t = 1$ , i.e., if  $(L, \pi^{-e_1}Q)$  is unimodular, then we have seen that .

$$Z(s) = q^{-e_1s}A_1.$$

On the other hand, by definition

$$\begin{aligned} Z_1 &= [1][0, 1] / [1, 1][0, 2][n_1, 2] \cdot q^{-e_1s} \{[n_1]q^{-s} + [n_1, 2] \\ &\quad - \chi(L_1)[0, 2]q^{-n_1/2}\}; \end{aligned}$$

and clearly they are equal. Therefore we shall assume that  $t \geq 2$  and apply an induction on  $t$ .

If for every integer  $j$  we put

$$\varphi(j) = \int_L |Q_1(x_1) + \sum_{i \geq 2} \pi^{e_i - e_1 - 2j} Q_i(x_i)|_K^s dx,$$

then we can write

$$Z(s) = \int_L \left| \sum_{1 \leq i \leq t} \pi^{e_i} Q_i(x_i) \right|_K^s dx = q^{-e_1s} \varphi(0).$$

We define an integer  $k$  as  $e_2 - e_1 = 2k$  or  $e_2 - e_1 = 2k + 1$  according as  $e_2 - e_1$  is even or odd, and for  $j < k$  we apply SPF to  $\varphi(j)$ . Then after a small computation we get

$$\varphi(j) = [n_1, 2]A_1 + q^{-n_1-2s}\varphi(j+1).$$

If we take  $j = 0, 1, \dots, k-1$  and eliminate  $\varphi(1), \dots, \varphi(k-1)$  from the resulting relations,

$$\varphi(0) = [n_1 k, 2k] A_1 + q^{-k(n_1 + 2s)} \varphi(k),$$

hence

$$(*) \quad Z(s) = [n_1 k, 2k] q^{-e_1 s} A_1 + q^{-n_1 k - (e_1 + 2k)s} \varphi(k).$$

We shall separate cases according as  $e_2 - e_1$  is even or odd.

Case 1:  $e_2 - e_1 = 2k$ .

If we put

$$Q'(x) = Q_1(x_1) + Q_2(x_2) + \sum_{i \geq 3} \pi^{e_i - e_2} Q_i(x_i),$$

then we get

$$\varphi(k) = \int_L |Q'(x)|_K^s dx.$$

Since  $(L_1 + L_2) + \dots + L_t$  is the Jordan decomposition of  $(L, Q')$ , we can apply an induction on  $t$  to  $\varphi(k)$ . In that way we get

$$\varphi(k) = A_{12} + q^{n_1(e_2 - e_1)/2 + e_2 s} \sum_{i \geq 3} Z_i.$$

By (\*) we have only to verify, therefore, that

$$\begin{aligned} Z_1 + Z_2 &= [n_1 k, 2k] q^{-e_1 s} A_1 + q^{-n_1 k - (e_1 + 2k)s} A_{12}, \text{ i.e.,} \\ Z_1 &= q^{-e_1 s} A_1, \quad Z_2 = q^{-n_1(e_2 - e_1)/2 - e_2 s} (A_{12} - A_1). \end{aligned}$$

We already know the first identity and we can verify the second identity by Lemma 2.

Case 2:  $e_2 - e_1 = 2k + 1$

In this case we have

$$\varphi(k) = \int_L |Q_1(x_1) + \sum_{i \geq 2} \pi^{e_i - e_2 + 1} Q_i(x_i)|_K^s dx.$$

If for every integer  $j$  we put

$$\begin{cases} \psi(2j) = \int_L |Q_1(x_1) + \pi Q_2(x_2) + \sum_{i \geq 3} \pi^{e_i - e_2 - 2j + 1} Q_i(x_i)|_K^s dx \\ \psi(2j + 1) = \int_L |Q_2(x_2) + \pi Q_1(x_1) + \sum_{i \geq 3} \pi^{e_i - e_2 - 2j} Q_i(x_i)|_K^s dx, \end{cases}$$

then  $\varphi(k) = \psi(0)$ . In the special case where  $t = 2$  we only have  $\psi(0)$  and  $\psi(1)$ . If we apply SPF to them, we get

$$\psi(0) = [n_1, 2] A_1 + q^{-n_1 - s} \psi(1), \quad \psi(1) = [n_2, 2] A_2 + q^{-n_2 - s} \psi(0).$$

By eliminating  $\psi(1)$  from these, we get

$$\varphi(k) = \{[n_1, 2]A_1 + [n_2, 2]q^{-n_1-s}A_2\}/[n_1 + n_2, 2].$$

By (\*) we have only to verify that

$$\begin{aligned} Z_1 + Z_2 &= [n_1 k, 2k]q^{-e_1 s}A_1 + q^{-n_1 k - (e_1 + 2k)s}\varphi(k), \text{ i.e.,} \\ Z_1 &= q^{-e_1 s}A_1, \quad Z_2 = q^{-n_1(e_2 - e_1 - 1)/2 - (e_2 - 1)s}(\varphi(k) - A_1). \end{aligned}$$

Again the second identity can be verified by Lemma 2.

In the general case where  $t \geq 3$  we apply SPF to  $\psi(2j)$  and  $\psi(2j+1)$  respectively for  $2j \leq e_3 - e_2$  and  $2j < e_3 - e_2$ ; then we get

$$\begin{aligned} \psi(2j) &= [n_1, 2]A_1 + q^{-n_1-s}\psi(2j+1) \\ \psi(2j+1) &= [n_2, 2]A_2 + q^{-n_2-s}\psi(2j+2). \end{aligned}$$

By using these relations alternatively, we can express  $\varphi(k) = \psi(0)$  by  $\psi(2j)$  and also by  $\psi(2j+1)$  both for  $2j \leq e_3 - e_2$ . We can take  $(e_3 - e_2)/2$  or  $(e_3 - e_2 - 1)/2$  as  $j$  according as  $e_3 - e_2$  is even or odd. In that way we get

$$\begin{aligned} \varphi(k) &= [(n_1 + n_2)(e_3 - e_2)/2, e_3 - e_2]\{[n_1, 2]A_1 \\ &\quad + [n_2, 2]q^{-n_1-s}A_2\}/[n_1 + n_2, 2] + q^{-(n_1 + n_2)(e_3 - e_2)/2 - (e_3 - e_2)s}\psi(e_3 - e_2) \end{aligned}$$

if  $e_3 - e_2$  is even and

$$\begin{aligned} \varphi(k) &= \{[(n_1 + n_2)(e_3 - e_2 + 1)/2, e_3 - e_2 + 1][n_1, 2]A_1 \\ &\quad + [(n_1 + n_2)(e_3 - e_2 - 1)/2, e_3 - e_2 - 1][n_2, 2]q^{-n_1-s}A_2\}/[n_1 + n_2, 2] \\ &\quad + q^{-(n_1 + n_2)(e_3 - e_2 - 1)/2 - n_1 - (e_3 - e_2)s}\psi(e_3 - e_2) \end{aligned}$$

if  $e_3 - e_2$  is odd. Again we shall separate cases.

Case 2.1  $e_3 - e_2$  even.

If we put

$$Q''(x) = Q_1(x_1) + \pi(Q_2(x_2) + Q_3(x_3)) + \sum_{i \geq 4} \pi^{e_i - e_3 + 1} Q_i(x_i),$$

then we get

$$\psi(e_3 - e_2) = \int_L |Q''(x)|_K^s dx.$$

Since  $L_1 + (L_2 + L_3) + \dots + L_t$  is the Jordan decomposition of  $(L, Q'')$ , by induction we get

$$\begin{aligned} & \cdot \{ \chi(L_1)[n_2 + n_3]q^{-n_1-s} + [n_1 + n_2 + n_3, 2]q^{-n_1/2} \\ & - \chi(L_2 + L_3)[n_1, 2]q^{-(n_1+n_2+n_3)/2} \}, \\ e &= n_1(e_3 - e_1 - 1)/2 + n_2(e_3 - e_2)/2 + (e_3 - 1)s. \end{aligned}$$

Therefore, as before, we have only to verify that

$$\begin{aligned} Z_2 &= q^{-n_1(e_2-e_1-1)/2-(e_2-1)s}/[n_1 + n_2, 2] \\ & \cdot \{ ([n_1, 2] - [n_1 + n_2, 2])A_1 + [n_2, 2]q^{-n_1-s}A_2 \}, \\ Z_3 &= q^{-e} \{ B - ([n_1, 2]A_1 + [n_2, 2]q^{-n_1-s}A_2)/[n_1 + n_2, 2] \}. \end{aligned}$$

The first identity follows from Lemma 2. As for the second identity, we eliminate  $A_2$  by Lemma 2; then we see that the coefficients of  $\chi(L_2)$ ,  $\chi(L_2 + L_3)$  on both sides are trivially equal and the coefficients of  $\chi(L_1)$  are equal by Lemma 3.

Case 2.2:  $e_3 - e_2$  odd

If we put

$$Q''(x) = Q_2(x_2) + \pi(Q_1(x_1) + Q_3(x_3)) + \sum_{i \geq 4} \pi^{e_i - e_3 + 1} Q_i(x_i),$$

then  $L_2 + (L_1 + L_3) + \dots + L_t$  becomes the Jordan decomposition of  $(L, Q'')$  and

$$\psi(e_3 - e_2) = \int_L |Q''(x)|_K^s dx.$$

Therefore if we apply the permutation of the subscripts 1 and 2 to the above expressions for  $B$ ,  $e$ , we get

$$\psi(e_3 - e_2) = B + q^e \sum_{i \geq 4} Z_i$$

with the new  $B$ ,  $e$ . Furthermore the identity for  $Z_2$  to be verified is the same as in the previous case and the one for  $Z_3$  becomes

$$Z_3 = q^{-e} \{ B - ([n_2, 2]A_2 + [n_1, 2]q^{-n_2-s}A_1)/[n_1 + n_2, 2] \}.$$

We need no new verification because the old  $Z_3$  and the new  $Z_3$  differ only by the permutation of the subscripts 1 and 2. The induction is thus complete and Theorem 1 is proved.

## 6. Remarks

We have shown in Theorem 1 that

$$\int_L \left| \sum_{1 \leq i \leq t} \pi^{e_i} Q_i(x_i) \right|_K^s dx = \sum_{1 \leq i \leq t} Z_i,$$

in which  $L$  is the direct sum of free  $O_K$ -submodules  $L_1, \dots, L_t$  of  $V$ ,  $(L_i, Q_i)$  is unimodular

$K(L_1 + \dots + L_i)$  such that  $L_1 + \dots + L_i$  is of measure 1, then

$$\int_{L_1 + \dots + L_i} \left| \sum_{j \leq i} \pi^{e_j} Q_j(x_j) \right|_K^s dx_1 \dots dx_i = \sum_{j \leq i} Z_j$$

for  $1 \leq i \leq t$ . This is the first remark.

The second remark is that in Theorem 1 the condition  $e_1 < \dots < e_t$  can be replaced by  $e_1 \leq \dots \leq e_t$ . This can be proved as follows: if  $e_i = e_{i+1}$ , we put

$$L_i^* = L_i + L_{i+1}, \quad n_i^* = n_i + n_{i+1}, \quad e_i^* = e_i$$

and define  $M_1^*, M_2^*, M_3^*, m_1^*, m_2^*$  relative to  $L_1 + \dots + L_i^*$ ; then  $M_1^* = M_1, M_2^* = M_2, m_1^* = m_1$ . Therefore on the RHS we get

$$\begin{aligned} Z_i + Z_{i+1} &= [1][0, 1]/[1, 1][m_1^*, 2][m_2^*, 2] \cdot q^{-1/2 \cdot \sum_{j < i} n_j(e_j^* - e_j) - e_i^* s} \\ &\quad \cdot \{ \chi(M_1^*)[n_1^*] q^{-m_1^* - s} + \chi(M_2^*)[m_2^*, 2] q^{-n_1^*/2} \\ &\quad - \chi(M_3^*)[m_1^*, 2] q^{-m_2^*/2} \}. \end{aligned}$$

In fact, this is an identity with  $\chi(M_1), \chi(M_2), \chi(M_3), \chi(M_3^*)$  as variables: The coefficients of  $\chi(M_2), \chi(M_3), \chi(M_3^*)$  on both sides are trivially equal and the coefficients of  $\chi(M_1)$  are equal by Lemma 3. Therefore we can shorten  $Z_1 + \dots + Z_t$  as many times as we have equalities in  $e_1 \leq \dots \leq e_t$  and the shortened expression is equal to  $Z(s)$  on the LHS by Theorem 1.

## 7. $Z(s)$ for quadratic polynomials

If  $f(x)$  is any quadratic polynomial on  $V$  with nondegenerate second degree part, we can eliminate the first degree part by translation in  $V$ . We shall discuss two inhomogeneous cases which often appear in applications.

**Theorem 2.** Let  $Q$  denote any nondegenerate quadratic form on  $V$ ,  $L$  a lattice in  $V$ , and  $L = L_1 + \dots + L_t$  the Jordan decomposition of  $(L, Q)$  as in Theorem 1; take any integer  $e \geq e_t, u$  from  $O_K^\times$ , and define a quadratic form  $Q^*$  on  $V + K$  as  $Q^*(x + y) = Q(x) + \pi^e u y^2$ ; and put

$$M_1 = \sum_{e_i \neq e} L_i, \quad M_2 = \sum_{e_i = e} L_i, \quad M_3 = M_2 + O_K.$$

Then

$$\begin{aligned} \int_L |Q(x) + \pi^e u|_K^s dx &= \int_L |Q(x)|_K^s dx \\ &\quad + [0, 1] |D(L, Q)|_K^{-1/2} q^{-(n/2 + s)e} / [1, 1][n, 2] \\ &\quad \cdot \{ \chi(M_1)[1] q^{-n-s} + \chi(M_2)[n+1, 2] q^{-n/2} \\ &\quad - \chi(M_3)[n, 2] q^{-(n+1)/2} \}, \end{aligned}$$

in which  $n = \dim(V)$  and  $\chi(M_3)$  is defined relative to  $(L + O_K, Q^*)$ .

*Proof.* If we denote by  $dy$  the Haar measure on  $K$  such that  $O_K$  is of measure 1, then by splitting  $O_K$  into  $O_K^\times$  and  $\pi O_K$  we get

$$\begin{aligned} \int_{L+O_K} |Q(x) + \pi^e u y^2|_K^s dx dy &= [1] \cdot \int_L |Q(x) + \pi^e u|_K^s dx + q^{-1} \\ &\quad \cdot \int_{L+O_K} |Q(x) + \pi^{e+2} u y^2|_K^s dx dy. \end{aligned}$$

We apply Theorem 1 to the two integrals over  $L + O_K$  by using the remarks in the previous section. Then by a small elementary computation we get the formula in the theorem.

## COROLLARY

*In the same notation as in Theorem 2,*

$$\begin{aligned} \int_{L+O_K} |Q(x) + \pi^e y|_K^s dx dy &= \int_L |Q(x)|_K^s dx \\ &\quad + [1][0, 1] |D(L, Q)|_K^{-1/2} q^{-(n/2+s)e} / [1, 1][n, 2] \\ &\quad \cdot \{\chi(M_1) q^{-n-s} + \chi(M_2) q^{-n/2}\}. \end{aligned}$$

*Proof.* By splitting  $O_K$  into  $\pi^{2k} O_K^\times$ ,  $\pi^{2k+1} O_K^\times$  and  $\{0\}$  for all  $k \geq 0$  we get

$$\begin{aligned} \int_{L+O_K} |Q(x) + \pi^e y|_K^s dx dy &= \sum_{k \geq 0} q^{-2k} \int_{O_K^\times} \left\{ \int_L |Q(x) + \pi^{e+2k} u|_K^s dx \right\} du \\ &\quad + \sum_{k \geq 0} q^{-2k-1} \int_{O_K^\times} \left\{ \int_L |Q(x) + \pi^{e+2k+1} u|_K^s dx \right\} du. \end{aligned}$$

We apply Theorem 2 to the two integrals over  $L$  and then use the fact that the integral of  $\chi(M_3)$  over  $O_K^\times$  is 0 in view of

$$\int_{O_K^\times} \chi(u) du = [1]/2 \cdot (1 - 1) = 0.$$

The rest is a small elementary computation.

## 8. An application

A typical form of difficult auxiliary integrals we have encountered is

$$\int_{O_K^\times \times O_K^\times} |y A(x) y + b(x)|_K^s dx dy,$$

in which  $A(x)$  is a nonsingular symmetric matrix of degree  $n$  with entries in  $O_K[x_1, \dots, x_m]$  and, up to a factor in  $O_K$ , certain powers of  $\det(A(x))$  and  $b(x)$  are equal; cf., e.g., [3], p. 218. In this last section we shall give a closed form to the



in which  $\text{Sym}_n$  denotes the space of symmetric matrices of degree  $n$  and  $e \geq 0$ .

We first state the Jordan decomposition theorem in a matrix form: We choose representatives  $1, \varepsilon$  of  $O_K^\times / (O_K^\times)^2$  so that  $\Sigma \chi(u)$  for  $u = 1, \varepsilon$  is 0; we let  $GL_n(O_K)$  act on  $\text{Sym}_n(O_K) \cap GL_n(K)$  as  $(g, x) \mapsto g \cdot x = g x {}^t g$ . Then every orbit has a unique representative of the form

$$h = \begin{pmatrix} \pi^{e_1} h_1 & & \\ & \ddots & \\ & & \pi^{e_t} h_t \end{pmatrix}, \quad h_i = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \\ & & & u_i \end{pmatrix},$$

in which  $u_i = 1, \varepsilon$  for every  $i$  and  $0 \leq e_1 < \dots < e_t$ . Furthermore if

$$\deg(h_i) = n_i, \quad d_i = (-1)^{n_i(n_i-1)/2} u_i$$

for  $1 \leq i \leq t$  and if  $\mu$  denotes the Haar measure on  $\text{Sym}_n(K)$  such that  $\text{Sym}_n(O_K)$  is of measure 1, then

$$\mu(GL_n(O_K) \cdot h) = \mu(GL_n) / \prod_{1 \leq i \leq t} \mu(O(h_i)) \cdot q^{-1/2 \cdot \sum_i n_i e_i (n_i + 2 \cdot \sum_{j > i} n_j + 1)},$$

in which  $\mu(GL_n)$ ,  $\mu(O(h_i))$  are the canonical volumes of  $GL_n(O_K)$ ,  $O(h_i)(O_K)$ :

$$\mu(GL_n) = \prod_{1 \leq j \leq n} [j],$$

$$\mu(O(h_i)) = 2 \cdot \prod_{1 \leq j < n_i/2} [2j] \cdot \begin{cases} 1 - \chi(d_i) q^{-n_i/2} & n_i \text{ even} \\ 1 & n_i \text{ odd} \end{cases}$$

for  $1 \leq i \leq t$ . The above formula must be well known; it can be proved by an induction on  $t$ . The following lemma can also be proved by an induction on  $t$ :

**Lemma 4.** Let  $s_i$  denote any complex number with  $\text{Re}(s_i) > 0$  and  $r_i = 0, 1$  for  $1 \leq i \leq t$ ; put

$$\Phi(r, s) = \sum_{k_i} q^{-(k_1 s_1 + \dots + k_t s_t)},$$

in which  $k_1, \dots, k_t$  are integers satisfying  $0 \leq r_1 + 2k_1 < \dots < r_t + 2k_t$ . Then the series is absolutely convergent and it represents

$$q^{-\sum_{1 \leq i \leq t} (i-1 + \sum_{1 \leq j \leq i} (r_j (r_j - 1)) s_i)} \prod_{1 \leq i \leq t} (1 - q^{-\sum_{i \leq j \leq t} s_j}).$$

We have separated a special case of Theorem 3 in the following lemma:

Lemma 5. We have

$$\int_{\text{Sym}_n(O_K) \times O_K^n} |^t yxy|_K^s dx dy = [1][n]/[1, 1][n, 2].$$

*Proof.* We put  $W_n = O_K^n - \pi O_K^n$  and split  $O_K^n$  into  $\pi^k W_n$  and  $\{0\}$  for all  $k \geq 0$ ; then we get

$$\int_{\text{Sym}_n(O_K) \times O_K^n} |^t yxy|_K^s dx dy = \sum_{k \geq 0} q^{-(n+2s)k} \int_{W_n} \left\{ \int_{\text{Sym}_n(O_K)} |^t yxy|_K^s dx \right\} dy.$$

If  $\eta = (10 \dots 0)$ , then every  $y$  in  $W_n$  can be written as  $g\eta$  for some  $g$  in  $GL_n(O_K)$  and the action of  $GL_n(O_K)$  on  $\text{Sym}_n(O_K)$  is measure preserving. Therefore the above integral over  $\text{Sym}_n(O_K)$  is  $[1]/[1, 1]$ , hence the RHS becomes  $[1]/[1, 1] \cdot [n]/[n, 2]$ .

**Theorem 3.** We take a partition  $n = n_1 + \dots + n_t$  of  $n$ , choose  $r_i = 0, 1$ , put

$$s_i = n_i \left( n + 2s + n_i + 2 \sum_{j>i} n_j \right)$$

for  $1 \leq i \leq t$ , and for a given  $e \geq 0$  we split  $\{1, 2, \dots, t\}$  into two subsets  $I, J$  as

$$I(\text{resp. } J) = \left\{ i; r_i \equiv (\text{resp. } \not\equiv) e + \sum_{1 \leq j \leq t} n_j r_j \pmod{2} \right\};$$

and finally we define  $A_S$  for  $S = I, J$  as

$$A_S = \prod_{i \in S} \delta_0(n_i) q^{-n_i/2} \left/ \prod_{1 \leq i \leq t} \prod_{1 \leq j \leq n_i/2} [2j] \right.,$$

in which  $\delta_r(m)$  for any integers  $r, m$  represents 1 or 0 according as  $m - r$  is even or odd. Then

$$\begin{aligned} & \int_{\text{Sym}_n(O_K) \times O_K^n} |^t yxy + (-1)^{n(n+1)/2} \pi^e \det(x)|_K^s dx dy = [1][n]/[1, 1][n, 2] \\ & + \prod_{1 \leq i \leq n} [i] \cdot [0, 1] q^{-(n/2+s)e}/[1, 1][n, 2] \\ & \cdot \sum_{n_i} \sum_{r_i} \{ [n+1, 2] A_I + ([1] q^{-n/2-s} - \delta_1(n) [n, 2] q^{-1/2}) A_J \} \\ & \cdot q^{-1/2 \cdot (n + \sum_j r_j s_j)} \Phi(r, s) \end{aligned}$$

with  $\Phi(r, s)$  as in Lemma 4.

*Proof.* We decompose the LHS according to the above-recalled splitting of  $\text{Sym}_n(O_K) \cap GL_n(K)$  and apply Theorem 2 to each integral. Then by using the expression for  $\mu(GL_n(O_K) \cdot h)$  and Lemma 5 we get

$$\text{LHS} = [1][n]/[1, 1][n, 2] + \mu(GL_n)[0, 1] q^{-(n/2+s)e}/[1, 1][n, 2]$$

$$\sum_{n_i, e_i, u_i} q^{-1/2 \cdot \sum_j s_j e_j} \left/ \prod_{1 \leq j \leq t} \mu(O(h_j)) \right. \\ \cdot \{ \chi(M_1)[1]q^{-n-s} + \chi(M_2)[n+1, 2]q^{-n/2} - \chi(M_3)[n, 2]q^{-(n+1)/2} \}.$$

Furthermore if  $e_i \equiv r_i \pmod 2$  with  $r_i = 0, 1$  for  $1 \leq i \leq t$ , then

$$\sum_{u_1, \dots, u_t = 1, \varepsilon} 1 \left/ \prod_{1 \leq j \leq t} \mu(O(h_j)) \right. \cdot \chi(M_k) = A_J, A_I, \delta_1(n) A_J$$

respectively for  $k = 1, 2, 3$ . Therefore the summation in  $e_i, u_i$  above becomes

$$\sum_{r_i} \{ [n+1, 2] A_I + ([1]q^{-n/2-s} - \delta_1(n)[n, 2]q^{-1/2}) A_J \} \cdot q^{-1/2 \cdot (n + \sum_j r_j s_j)} \Phi(r, s).$$

This completes the proof.

We might mention that in Theorem 3 if we replace  $(-1)^{n(n+1)/2}$  by  $\pm 1$ , then we have only to insert the factor  $\chi(\pm (-1)^{n(n+1)/2})$  after  $\delta_1(n)$ . The special case of Theorem 3 for  $n=4$ ,  $e=1$  appears as the lowest codimensional new partial integral in the computation of  $Z(s)$  for the degree 8 invariant of  $\text{Spin}_{14}$ . In that case the theorem gives

$$\int_{\text{Sym}_4(O_K) \times O_K^*} |^t yxy + \pi \det(x)|_K^s dx dy = [1] \{ [4] + [3]q^{-4-s} \\ - [7]q^{-5-2s} \} / [1, 1][5, 2][7, 2].$$

## Acknowledgement

This work was partially supported by the National Science Foundation.

## References

- [1] Bourbaki N, Algèbre, Chapitre V: Corps Commutatifs, Hermann (1950)
- [2] Igusa J, Complex powers and asymptotic expansions. II, *Crelles J. Math.*, **278/279** (1975) 307–321; or Forms of Higher Degree, *Tata Inst. Lect. Notes* **59**, Springer (1978)
- [3] Igusa J, On the arithmetic of a singular invariant, *Am. J. Math.* **110** (1988), 197–233
- [4] Igusa J, A stationary phase formula for  $p$ -adic integrals and its applications, *Algebraic geometry and its applications*, Springer (1993) 193–212
- [5] O'Meara O T, Introduction to quadratic forms, *Grundl. Math. Wiss.*, **117**, Springer (1971)



# Vector bundles as direct images of line bundles

A HIRSCHOWITZ and M S NARASIMHAN\*

Université de Nice Sophia-Antipolis, Parc Valrose, 06108 Nice Cedex 2, France

\*International Centre for Theoretical Physics, P.O. Box 586, 34100 Trieste, Italy

Dedicated to the memory of Professor K G Ramanathan

**Abstract.** Let  $X$  be a smooth irreducible projective variety over an algebraically closed field  $K$  and  $E$  a vector bundle on  $X$ . We prove that, if  $\dim X \geq 1$ , there exist a smooth irreducible projective variety  $Z$  over  $K$ , a surjective separable morphism  $f: Z \rightarrow X$  which is finite outside an algebraic subset of codimension  $\geq 3$  in  $X$  and a line bundle  $L$  on  $X$  such that the direct image of  $L$  by  $f$  is isomorphic to  $E$ . When  $X$  is a curve, we show that  $Z, f, L$  can be so chosen that  $f$  is finite and the canonical map

$$H^1(Z, \mathcal{O}) \rightarrow H^1(X, \text{End } E)$$

is surjective.

**Keywords.** Projective variety; algebraic vector bundle; line bundle; direct image; finite morphism.

## 1. Introduction

Let  $X$  be a smooth irreducible projective variety over an algebraically closed field and  $E$  a vector bundle on  $X$ . We prove in this paper first that, if  $\dim X \geq 1$ ,  $E$  is the direct image of a line bundle  $L$  on a smooth irreducible projective variety  $Z$  by a morphism  $f: Z \rightarrow X$  which is finite outside an algebraic subset of codimension  $\geq 3$  in  $X$ . Moreover one can choose the morphism  $f$  to be separable and to have the property that all higher direct images of  $L$  by  $f$  are zero [Theorem 4.2].

In particular if  $\dim X \leq 2$  the morphism  $f$  may be chosen to be finite. In the case of surfaces this result has been proved by R.L.E. Schwarzenberger for rank two vector bundles [5, Theorem 3]. We also give an example of a vector bundle on  $\mathbf{P}_3$  which cannot be obtained as the direct image of a line bundle on a smooth variety by a finite morphism.

In the second part of the paper we consider the case when  $X$  is a curve. We prove in this case that  $Z, L$  and  $f$  can be so chosen that the canonical homomorphism (see 5.1)

$$H^1(Z, \mathcal{O}_Z) \rightarrow H^1(X, \text{End } E)$$

is surjective (Theorem 6.4). This result was proved for a "very stable" vector bundle  $E$  by Beauville-Narasimhan-Ramanan in the case of a curve over  $\mathbf{C}$  by using the Hitchin map [3]. (For the significance of this result see Remark 6.5).

Let  $\pi: \mathbf{P}(E) \rightarrow X$  be the projective bundle associated to  $E$ . The variety  $Z$  is constructed as the subscheme (of  $\mathbf{P}(E)$ ) of zeros of a generic section of the tangent bundle along the fibres of  $\pi$  twisted by a suitable ample line bundle on  $X$  pulled up to  $\mathbf{P}(E)$ ; the line bundle  $L$  is simply taken to be the restriction of  $\mathcal{O}_{\mathbf{P}(E)}(1)$  to  $Z$ .

The scheme  $Z$  is essentially the scheme of 'eigenstates' of a generic twisted endomorphism of  $E$ . However, in general,  $Z$  is not the spectral variety of the twisted endomorphism; the canonical map from the spectral variety into  $X$  is always a finite morphism.

## 2. Sections of the tangent bundle of a projective space

Let  $V$  be a finite dimensional vector space of dimension  $\geq 2$  over an algebraically closed field  $K$  and  $\mathbf{P} = \mathbf{P}(V)$  the projective space of hyperplanes in  $V$ . We have the exact sequences of vector bundles on  $\mathbf{P}(E)$ :

$$0 \rightarrow \Omega^1(1) \rightarrow V_{\mathbf{P}} \rightarrow \mathcal{O}(1) \rightarrow 0$$

and

$$0 \rightarrow \mathcal{O}_{\mathbf{P}} \rightarrow V_{\mathbf{P}}^* \otimes \mathcal{O}(1) \rightarrow \Theta \rightarrow 0,$$

where  $\Theta$  denotes the tangent bundle of  $\mathbf{P}$  and  $V^*$  the dual of  $V$ . We obtain an exact sequence of vector spaces

$$0 \rightarrow K \rightarrow V^* \otimes V \rightarrow H^0(\mathbf{P}, \Theta) \rightarrow 0.$$

If  $\text{End}^0(V) := \text{End}(V)/(\text{Scalar endomorphisms})$  we have  $\text{End}^0(V) = H^0(\mathbf{P}, \Theta)$ .

If an endomorphism  $T$  of  $V$  leaves a hyperplane  $\xi$  invariant,  $T$  induces an endomorphism of the one dimensional space  $V/\xi$ . The subspace  $(V/\xi)^*$  of  $V^*$  is an eigenspace of the transpose of  $T$ . If  $s_T$  is the section of  $\Theta$  defined by  $T$ , we have  $s_T(\xi) = 0$  if and only if  $T(\xi) \subset \xi$ . Thus we can view the subscheme  $Z = Z_T$  of zeros of  $s_T$  as the scheme of "eigenstates" of  $T$ . Moreover the "eigenvalue" of  $T$  is a section of  $\mathcal{O}_Z$ ; in fact it is the section of  $\mathcal{O}_Z$  corresponding to the morphism  $\mathcal{O}_Z(1) \rightarrow \mathcal{O}_Z(1)$  induced by  $T$  from:

$$\begin{array}{c} 0 \rightarrow \Omega^1(1) \rightarrow V_{\mathbf{P}} \rightarrow \mathcal{O}(1) \rightarrow 0 \\ \quad \quad \quad T \downarrow \\ \quad \quad \quad V_{\mathbf{P}} \rightarrow \mathcal{O}(1) \rightarrow 0. \end{array}$$

Observe that the scheme  $Z_T$  has dimension  $\geq 1$  if and only if the transpose of  $T$  has an eigenspace of dimension  $\geq 2$  corresponding to some eigenvalue.

### PROPOSITION 2.1

*Consider the exact sequence of vector bundles*

$$0 \rightarrow F \rightarrow \text{End}^0(V)_{\mathbf{P}} \rightarrow \Theta \rightarrow 0$$

*( $F$  being defined as the kernel of homomorphism  $\text{End}^0(V)_{\mathbf{P}} \rightarrow \Theta$ ). Let  $p: F \rightarrow \text{End}^0(V)$  be the restriction to  $F$  of the projection  $\text{End}^0(V) \times \mathbf{P} \rightarrow \text{End}^0(V)$ . Then there exists an open subset  $\Omega$  in  $\text{End}^0(V)$ , whose complement is of codimension  $\geq 3$ , such that the morphism  $p: p^{-1}(\Omega) \rightarrow \Omega$  is finite.*

*Proof.* Consider the commutative diagram

$$\begin{array}{ccccccc}
 & 0 & & 0 & & & \\
 & \downarrow & & \downarrow & & & \\
 & K_{\mathbf{P}} & & K_{\mathbf{P}} & & & \\
 & \downarrow & & \downarrow & & & \\
 0 & \rightarrow & F^1 & \rightarrow & \text{End}(V)_{\mathbf{P}} & \rightarrow & \Theta \rightarrow 0 \\
 & & \downarrow & & \downarrow & & \parallel \\
 0 & \rightarrow & F & \rightarrow & \text{End}^0(V)_{\mathbf{P}} & \rightarrow & \Theta \rightarrow 0 \\
 & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 
 \end{array}$$

Let  $q: F^1 \rightarrow \text{End}(V)$  be the projection. We shall show that there exists an open set  $U$  of  $\text{End}(V)$  which is saturated for the map  $\text{End}(V) \rightarrow \text{End}^0(V)$  and whose complement is of codimension  $\geq 3$  such that the morphism  $q: q^{-1}(U) \rightarrow U$  is finite. This will prove the proposition.

For each subspace  $W$  of dimension  $k \geq 2$  of  $V^*$ , consider the subspace of  $\text{End}(V^*)$  consisting of those endomorphisms whose restriction to  $W$  is a scalar endomorphism of  $W$ . The dimension of this space is  $1 + (r-k)^2 + k(r-k)$ . Varying  $W$  over the Grassmannian  $G(r, k)$  we get a vector bundle  $W(r, k)$  over  $G(r, k)$  and the dimension of the total space of this bundle is

$$1 + (r-k)^2 + k(r-k) + k(r-k) = 1 + r^2 - k^2.$$

We have a natural morphism  $\pi_k: W(r, k) \rightarrow \text{End}(V)$  which maps an endomorphism to its transpose. If  $S_k := \pi_k(W(r, k))$ , we have  $\dim S_k \leq (1 + r^2 - k^2)$  and  $\text{codim } S_k \geq k^2 - 1 \geq 3$ . Let  $S = \bigcup_{2 \leq k \leq r} S_k$  and  $U = \text{End}(V) - S$ . We have  $\text{codim } S \geq 3$  and  $S$  is saturated for the map  $\text{End}(V) \rightarrow \text{End}^0(V)$ . The fibres of  $q: q^{-1}(U) \rightarrow U$  are finite and  $q$  is proper. Hence  $q$  is a finite morphism.

### 3. Sections of the (twisted) relative tangent bundle of a projective bundle

Let  $E$  be a vector bundle of rank  $r \geq 2$  on a smooth irreducible projective variety  $X$  of dimension  $\geq 1$  over  $K$ . Let  $\pi: \mathbf{P}(E) \rightarrow X$  be the associated projective bundle. We have the exact sequences on  $\mathbf{P}(E)$ :

$$0 \rightarrow \Omega_{\pi}^1(1) \rightarrow \pi^*(E) \rightarrow \mathcal{O}_{\mathbf{P}(E)}(1) \rightarrow 0$$

and

$$0 \rightarrow \mathcal{O}_{\mathbf{P}(E)} \rightarrow \pi^*(E^*) \otimes \mathcal{O}(1) \rightarrow \Theta_{\pi} \rightarrow 0$$

where  $\Theta_{\pi}$  (resp.  $\Omega_{\pi}^1$ ) denotes the relative tangent (resp. cotangent) bundle along the fibres of  $\pi$ . Let  $\text{End}^0(E)$  denote the vector bundle  $\text{End}(E)/\mathcal{O}_X$ . We have an exact sequence of vector bundles on  $\mathbf{P}(E)$ :

$$0 \rightarrow F \rightarrow \pi^*(\text{End}^0(E)) \rightarrow \Theta_{\pi} \rightarrow 0.$$

Let  $M$  be a line bundle on  $X$ . We obtain the exact sequence:

$$0 \rightarrow F \otimes \pi^*(M) \rightarrow \pi^*(\text{End}^0(E) \otimes M) \rightarrow \Theta_\pi \otimes \pi^*(M) \rightarrow 0$$

### PROPOSITION 3.1

Let  $p: F \otimes \pi^*(M) \rightarrow \text{End}^0(E) \otimes M$  be the canonical morphism (of total spaces of geometric vector bundles over  $\mathbf{P}(E)$  and  $X$  respectively). Then there exists an open subset  $\Omega$  of  $\text{End}^0(E) \otimes M$  whose complement is of codimension  $\geq 3$  such that the morphism  $p: p^{-1}(\Omega) \rightarrow \Omega$  is finite.

*Proof.* This follows from Proposition 2.1.

### PROPOSITION 3.2

There exists an ample line bundle  $M$  on  $X$  such that a generic section  $s$  of  $\Theta_\pi \otimes \pi^*(M)$  (i.e. for  $s$  in a non-empty open subset of  $H^0(\mathbf{P}(E), \Theta_\pi \otimes \pi^*(M))$ ) satisfies the following conditions:

- (a) The scheme  $Z$  of zeros of  $s$  is smooth and irreducible.
- (b) The morphism  $\pi|_Z: Z \rightarrow X$  is surjective and separable.
- (c) There exists a closed subset  $S$  of  $X$  of codimension  $\geq 3$  such that the morphism

$$\pi: Z \setminus \pi^{-1}(S) \rightarrow X \setminus S$$

is finite.

*Proof.* Let  $\xi$  be an ample line bundle on  $X$ . Then the line bundle  $\pi^*(\xi^k) \otimes \mathcal{O}(1)$  is very ample on  $\mathbf{P}(E)$  for  $k \geq k_0$ . [4, II, Prop. 7.10, p. 161]. We may also assume that for  $k \geq k_0$ , the bundle  $\xi^k \otimes E^*$  is generated by its sections. Let  $M = \xi^{2k}$ . Since the sections of the very ample line bundle  $\pi^*(\xi^k) \otimes \mathcal{O}(1)$  generate its first order jet bundle and  $\pi^*(\xi^k \otimes E^*)$  is generated by its sections, we see that the sections of  $\pi^*(M \otimes E^*) \otimes \mathcal{O}(1)$  generate its first order jet bundle [7, Lemma 5]. Since  $\Theta_\pi \otimes \pi^*(M)$  is a quotient bundle of  $\pi^*(M \otimes E^*) \otimes \mathcal{O}(1)$ , the sections of  $\Theta_\pi \otimes \pi^*(M)$  generate its first order jet bundle. Now by [7, Theorem 1] the zero scheme  $Z$  of a generic section  $s$  of  $\Theta_\pi \otimes \pi^*(M)$  is smooth. We will prove in the next proposition (Proposition 4.1) that  $Z$  is irreducible.

Let  $x_0 \in X$  and  $S := \pi^{-1}(x_0)$  the fibre over  $x_0$ . Let  $W$  be the image of the homomorphism

$$H^0(\mathbf{P}(E), \Theta_\pi \otimes \pi^*(M)) \rightarrow H^0(S, \Theta_\pi \otimes \pi^*(M)|_S).$$

(We may even assume that  $W = H^0(S, \Theta_\pi \otimes \pi^*(M)|_S)$ , by choosing  $k$  large enough.) Then the first order jets of elements of  $W$  generate the first order jet bundle of  $\Theta_\pi \otimes \pi^*(M)|_S$ ; hence, again by [7, Theorem 1], for a generic element  $\sigma$  of  $W$ , the zero subscheme (of  $S$ ) defined by  $\sigma$  is smooth. Thus we see that for a generic section  $s$  of  $\Theta_\pi \otimes \pi^*(M)$  the zero scheme  $Z$  is smooth and irreducible and  $Z$  intersects  $S$  transversally. It follows that there exists a point  $z_0 \in Z \cap S$  such that the differential of  $\pi|_Z$  at  $z_0$  is an isomorphism. This proves that  $\pi|_Z: Z \rightarrow X$  is surjective and (assuming  $Z$  to be irreducible) separable. (Observe that  $Z$  intersects every fibre  $\pi^{-1}(x)$ ,  $x \in$



for otherwise the tangent bundle of the projective space  $\pi^{-1}(x)$  would contain a trivial line bundle.)

$$\begin{aligned}\text{Let } \Sigma &:= H^0(\mathbf{P}(E), \pi^*(M) \otimes \Theta_\pi) \\ &= H^0(X, M \otimes \pi_*(\Theta_\pi)) \\ &= H^0(X, M \otimes \text{End}^0(E)).\end{aligned}$$

Consider the morphisms

$$\begin{array}{c} \varphi: \Sigma \times X \xrightarrow{\varphi} M \otimes \text{End}^0(E) \\ \downarrow p_X \\ X. \end{array}$$

where the evaluation map  $\varphi$  is a smooth morphism, being a surjection of vector bundles. Let  $\Omega$  be the open subset of  $M \otimes \text{End}^0(E)$  defined in Proposition 3.1 and  $N$  its complement. Then

$$\dim \varphi^{-1}(N) \leq \dim \Sigma + \dim X - 3.$$

By considering  $p_X: \varphi^{-1}(N) \rightarrow X$  we see that for a generic section  $s$  of  $M \otimes \text{End}^0(E)$  we have

$$\dim(\varphi^{-1}(N) \cap \{s \times X\}) \leq (\dim X) - 3.$$

Let  $S \subset X$  be defined to be  $p_X(\varphi^{-1}(N) \cap \{s \times X\})$ . Then  $\dim S \leq (\dim X) - 3$  and

$$\pi|_Z: Z \setminus \pi^{-1}(S) \rightarrow X \setminus S$$

is finite.

Thus for a generic section  $s$  of  $\Theta_\pi \otimes \pi^*(M)$  all the conditions a), b) and c) are satisfied.

#### 4. Koszul resolution of the zero scheme $Z$

*Proof of Theorem 4.2*

##### PROPOSITION 4.1

Let  $s$  be a section of  $\Theta_\pi \otimes \pi^*(M)$  over  $\mathbf{P}(E)$  such that the zero scheme  $Z$  of  $s$  is smooth. We then have

- a)  $\pi_*(\mathcal{O}_Z \otimes \mathcal{O}_{\mathbf{P}(E)}(1)) \simeq E$  and  $R^i \pi_*(\mathcal{O}_Z(1)) = 0$  for  $i \geq 1$ .
- b)  $\pi_*(\mathcal{O}_Z)$  has a filtration  
 $0 = F_0 \subset F_1 \subset F_2 \subset \dots \subset F_i \subset \dots \subset F_r = \pi_*(\mathcal{O}_Z)$   
such that  $F_i/F_{i-1} \simeq M^{-(i-1)} (= (M^*)^{\otimes (i-1)})$  for  $1 \leq i \leq r$  (In particular  $F_1 \simeq \mathcal{O}_Z$ ).
- c)  $Z$  is irreducible (if  $\dim X \geq 1$  and  $M$  is ample).

*Proof.* Using our assumption on  $Z$ , we have a Koszul resolution for  $\mathcal{O}_Z$  on  $\mathbf{P}(E)$ : [1, Ch I, Lemma 4.2 and Ch. III, Propositions 4.10 and 4.11]

$$(A) \quad 0 \rightarrow \bigwedge^{r-1} (\Omega_\pi^1 \otimes \pi^*(M^*)) \rightarrow \dots \rightarrow \bigwedge^2 (\Omega_\pi^1 \otimes \pi^*(M^*)) \rightarrow \Omega_\pi^1 \otimes \pi^*(M^*) \rightarrow \mathcal{O}_{\mathbf{P}(E)} \rightarrow \mathcal{O}_Z \rightarrow 0$$

Lemma 5. We have

$$\int_{\text{Sym}_n(O_K) \times O_K^n} |yxy|_K^s dx dy = [1][n]/[1, 1][n, 2].$$

*Proof.* We put  $W_n = O_K^n - \pi O_K^n$  and split  $O_K^n$  into  $\pi^k W_n$  and  $\{0\}$  for all  $k \geq 0$ ; then we get

$$\int_{\text{Sym}_n(O_K) \times O_K^n} |yxy|_K^s dx dy = \sum_{k \geq 0} q^{-(n+2s)k} \int_{W_n} \left\{ \int_{\text{Sym}_n(O_K)} |yxy|_K^s dx \right\} dy.$$

If  $\eta = (10 \dots 0)$ , then every  $y$  in  $W_n$  can be written as  $g\eta$  for some  $g$  in  $GL_n(O_K)$  and the action of  $GL_n(O_K)$  on  $\text{Sym}_n(O_K)$  is measure preserving. Therefore the above integral over  $\text{Sym}_n(O_K)$  is  $[1]/[1, 1]$ , hence the RHS becomes  $[1]/[1, 1] \cdot [n]/[n, 2]$ .

**Theorem 3.** We take a partition  $n = n_1 + \dots + n_t$  of  $n$ , choose  $r_i = 0, 1$ , put

$$s_i = n_i \left( n + 2s + n_i + 2 \sum_{j > i} n_j \right)$$

for  $1 \leq i \leq t$ , and for a given  $e \geq 0$  we split  $\{1, 2, \dots, t\}$  into two subsets  $I, J$  as

$$I(\text{resp. } J) = \left\{ i; r_i \equiv (\text{resp. } \not\equiv) e + \sum_{1 \leq j \leq t} n_j r_j \pmod{2} \right\};$$

and finally we define  $A_S$  for  $S = I, J$  as

$$A_S = \prod_{i \in S} \delta_0(n_i) q^{-n_i/2} \left/ \prod_{1 \leq i \leq t} \prod_{1 \leq j \leq n_i/2} [2j] \right.,$$

in which  $\delta_r(m)$  for any integers  $r, m$  represents 1 or 0 according as  $m - r$  is even or odd. Then

$$\begin{aligned} & \int_{\text{Sym}_n(O_K) \times O_K^n} |yxy + (-1)^{n(n+1)/2} \pi^e \det(x)|_K^s dx dy = [1][n]/[1, 1][n, 2] \\ & + \prod_{1 \leq i \leq n} [i] \cdot [0, 1] q^{-(n/2+s)e} / [1, 1][n, 2] \\ & \cdot \sum_{n_i} \sum_{r_i} \{ [n+1, 2] A_I + ([1] q^{-n/2-s} - \delta_1(n) [n, 2] q^{-1/2}) A_J \} \\ & \cdot q^{-1/2 \cdot (n + \sum_j r_j s_j)} \Phi(r, s) \end{aligned}$$

with  $\Phi(r, s)$  as in Lemma 4.

*Proof.* We decompose the LHS according to the above-recalled splitting of  $\text{Sym}_n(O_K) \cap GL_n(K)$  and apply Theorem 2 to each integral. Then by using the expression for  $\mu(GL_n(O_K) \cdot h)$  and Lemma 5 we get

$$\text{LHS} = [1][n]/[1, 1][n, 2] + \mu(GL_n)[0, 1] q^{-(n/2+s)e} / [1, 1][n, 2]$$

$$\sum_{n_i, e_i, u_i} q^{-1/2 \cdot \sum_j s_j e_j} \left/ \prod_{1 \leq j \leq t} \mu(O(h_j)) \right. \\ \cdot \{ \chi(M_1)[1]q^{-n-s} + \chi(M_2)[n+1, 2]q^{-n/2} - \chi(M_3)[n, 2]q^{-(n+1)/2} \}.$$

Furthermore if  $e_i \equiv r_i \pmod{2}$  with  $r_i = 0, 1$  for  $1 \leq i \leq t$ , then

$$\sum_{u_1, \dots, u_t = 1, \varepsilon} 1 \left/ \prod_{1 \leq j \leq t} \mu(O(h_j)) \right. \cdot \chi(M_k) = A_J, A_I, \delta_1(n) A_J$$

respectively for  $k = 1, 2, 3$ . Therefore the summation in  $e_i, u_i$  above becomes

$$\sum_{r_i} \{ [n+1, 2] A_I + ([1]q^{-n/2-s} - \delta_1(n)[n, 2]q^{-1/2}) A_J \} \cdot q^{-1/2 \cdot (n + \sum_j r_j s_j)} \Phi(r, s).$$

This completes the proof.

We might mention that in Theorem 3 if we replace  $(-1)^{n(n+1)/2}$  by  $\pm 1$ , then we have only to insert the factor  $\chi(\pm (-1)^{n(n+1)/2})$  after  $\delta_1(n)$ . The special case of Theorem 3 for  $n=4$ ,  $e=1$  appears as the lowest codimensional new partial integral in the computation of  $Z(s)$  for the degree 8 invariant of  $\text{Spin}_4$ . In that case the theorem gives

$$\int_{\text{Sym}_4(O_K) \times O_K^4} |yxy + \pi \det(x)|_K^s dx dy = [1] \{ [4] + [3]q^{-4-s} \\ - [7]q^{-5-2s} \} / [1, 1][5, 2][7, 2].$$

## Acknowledgement

This work was partially supported by the National Science Foundation.

## References

- [1] Bourbaki N, Algèbre, Chapitre V: Corps Commutatifs, Hermann (1950)
- [2] Igusa J, Complex powers and asymptotic expansions. II, *Crelles J. Math.*, **278/279** (1975) 307–321; or Forms of Higher Degree, *Tata Inst. Lect. Notes* **59**, Springer (1978)
- [3] Igusa J, On the arithmetic of a singular invariant, *Am. J. Math.* **110** (1988), 197–233
- [4] Igusa J, A stationary phase formula for  $p$ -adic integrals and its applications, *Algebraic geometry and its applications*, Springer (1993) 193–212
- [5] O'Meara O T, Introduction to quadratic forms, *Grundl. Math. Wiss.*, **117**, Springer (1971)



# Vector bundles as direct images of line bundles

A HIRSCHOWITZ and M S NARASIMHAN\*

Université de Nice Sophia-Antipolis, Parc Valrose, 06108 Nice Cedex 2, France

\*International Centre for Theoretical Physics, P.O. Box 586, 34100 Trieste, Italy

Dedicated to the memory of Professor K G Ramanathan

**Abstract.** Let  $X$  be a smooth irreducible projective variety over an algebraically closed field  $K$  and  $E$  a vector bundle on  $X$ . We prove that, if  $\dim X \geq 1$ , there exist a smooth irreducible projective variety  $Z$  over  $K$ , a surjective separable morphism  $f: Z \rightarrow X$  which is finite outside an algebraic subset of codimension  $\geq 3$  in  $X$  and a line bundle  $L$  on  $X$  such that the direct image of  $L$  by  $f$  is isomorphic to  $E$ . When  $X$  is a curve, we show that  $Z, f, L$  can be so chosen that  $f$  is finite and the canonical map

$$H^1(Z, \mathcal{O}) \rightarrow H^1(X, \text{End } E)$$

is surjective.

**Keywords.** Projective variety; algebraic vector bundle; line bundle; direct image; finite morphism.

## 1. Introduction

Let  $X$  be a smooth irreducible projective variety over an algebraically closed field and  $E$  a vector bundle on  $X$ . We prove in this paper first that, if  $\dim X \geq 1$ ,  $E$  is the direct image of a line bundle  $L$  on a smooth irreducible projective variety  $Z$  by a morphism  $f: Z \rightarrow X$  which is finite outside an algebraic subset of codimension  $\geq 3$  in  $X$ . Moreover one can choose the morphism  $f$  to be separable and to have the property that all higher direct images of  $L$  by  $f$  are zero [Theorem 4.2].

In particular if  $\dim X \leq 2$  the morphism  $f$  may be chosen to be finite. In the case of surfaces this result has been proved by R.L.E. Schwarzenberger for rank two vector bundles [5, Theorem 3]. We also give an example of a vector bundle on  $\mathbf{P}_3$  which cannot be obtained as the direct image of a line bundle on a smooth variety by a finite morphism.

In the second part of the paper we consider the case when  $X$  is a curve. We prove in this case that  $Z, L$  and  $f$  can be so chosen that the canonical homomorphism (see 5.1)

$$H^1(Z, \mathcal{O}_Z) \rightarrow H^1(X, \text{End } E)$$

is surjective (Theorem 6.4). This result was proved for a "very stable" vector bundle  $E$  by Beauville-Narasimhan-Ramanan in the case of a curve over  $\mathbf{C}$  by using the Hitchin map [3]. (For the significance of this result see Remark 6.5).

Let  $\pi: \mathbf{P}(E) \rightarrow X$  be the projective bundle associated to  $E$ . The variety  $Z$  is constructed as the subscheme (of  $\mathbf{P}(E)$ ) of zeros of a generic section of the tangent bundle along the fibres of  $\pi$  twisted by a suitable ample line bundle on  $X$  pulled up to  $\mathbf{P}(E)$ ; the line bundle  $L$  is simply taken to be the restriction of  $\mathcal{O}_{\mathbf{P}(E)}(1)$  to  $Z$ .

The scheme  $Z$  is essentially the scheme of 'eigenstates' of a generic twisted endomorphism of  $E$ . However, in general,  $Z$  is not the spectral variety of the twisted endomorphism; the canonical map from the spectral variety into  $X$  is always a finite morphism.

## 2. Sections of the tangent bundle of a projective space

Let  $V$  be a finite dimensional vector space of dimension  $\geq 2$  over an algebraically closed field  $K$  and  $\mathbf{P} = \mathbf{P}(V)$  the projective space of hyperplanes in  $V$ . We have the exact sequences of vector bundles on  $\mathbf{P}(E)$ :

$$0 \rightarrow \Omega^1(1) \rightarrow V_{\mathbf{P}} \rightarrow \mathcal{O}(1) \rightarrow 0$$

and

$$0 \rightarrow \mathcal{O}_{\mathbf{P}} \rightarrow V_{\mathbf{P}}^* \otimes \mathcal{O}(1) \rightarrow \Theta \rightarrow 0,$$

where  $\Theta$  denotes the tangent bundle of  $\mathbf{P}$  and  $V^*$  the dual of  $V$ . We obtain an exact sequence of vector spaces

$$0 \rightarrow K \rightarrow V^* \otimes V \rightarrow H^0(\mathbf{P}, \Theta) \rightarrow 0.$$

If  $\text{End}^0(V) := \text{End}(V)/(\text{Scalar endomorphisms})$  we have  $\text{End}^0(V) = H^0(\mathbf{P}, \Theta)$ .

If an endomorphism  $T$  of  $V$  leaves a hyperplane  $\xi$  invariant,  $T$  induces an endomorphism of the one dimensional space  $V/\xi$ . The subspace  $(V/\xi)^*$  of  $V^*$  is an eigenspace of the transpose of  $T$ . If  $s_T$  is the section of  $\Theta$  defined by  $T$ , we have  $s_T(\xi) = 0$  if and only if  $T(\xi) \subset \xi$ . Thus we can view the subscheme  $Z = Z_T$  of zeros of  $s_T$  as the scheme of "eigenstates" of  $T$ . Moreover the "eigenvalue" of  $T$  is a section of  $\mathcal{O}_Z$ ; in fact it is the section of  $\mathcal{O}_Z$  corresponding to the morphism  $\mathcal{O}_Z(1) \rightarrow \mathcal{O}_Z(1)$  induced by  $T$  from:

$$\begin{array}{c} 0 \rightarrow \Omega^1(1) \rightarrow V_{\mathbf{P}} \rightarrow \mathcal{O}(1) \rightarrow 0 \\ \quad \quad \quad T \downarrow \\ \quad \quad \quad V_{\mathbf{P}} \rightarrow \mathcal{O}(1) \rightarrow 0. \end{array}$$

Observe that the scheme  $Z_T$  has dimension  $\geq 1$  if and only if the transpose of  $T$  has an eigenspace of dimension  $\geq 2$  corresponding to some eigenvalue.

### PROPOSITION 2.1

*Consider the exact sequence of vector bundles*

$$0 \rightarrow F \rightarrow \text{End}^0(V)_{\mathbf{P}} \rightarrow \Theta \rightarrow 0$$

( $F$  being defined as the kernel of homomorphism  $\text{End}^0(V)_{\mathbf{P}} \rightarrow \Theta$ ). Let  $p: F \rightarrow \text{End}^0(V)$  be the restriction to  $F$  of the projection  $\text{End}^0(V) \times \mathbf{P} \rightarrow \text{End}^0(V)$ . Then there exists an open subset  $\Omega$  in  $\text{End}^0(V)$ , whose complement is of codimension  $\geq 3$ , such that the morphism  $p: p^{-1}(\Omega) \rightarrow \Omega$  is finite.

*Proof.* Consider the commutative diagram

$$\begin{array}{ccccccc} & & 0 & & 0 & & \\ & & \downarrow & & \downarrow & & \\ & & K_P & & K_P & & \\ & & \downarrow & & \downarrow & & \\ 0 & \rightarrow & F^1 & \rightarrow & \text{End}(V)_P & \rightarrow & \Theta \rightarrow 0 \\ & & \downarrow & & \downarrow & & \parallel \\ 0 & \rightarrow & F & \rightarrow & \text{End}^0(V)_P & \rightarrow & \Theta \rightarrow 0 \\ & & \downarrow & & \downarrow & & \\ & & 0 & & 0 & & \end{array}$$

Let  $q:F^1 \rightarrow \text{End}(V)$  be the projection. We shall show that there exists an open set  $U$  of  $\text{End}(V)$  which is saturated for the map  $\text{End}(V) \rightarrow \text{End}^0(V)$  and whose complement is of codimension  $\geq 3$  such that the morphism  $q:q^{-1}(U) \rightarrow U$  is finite. This will prove the proposition.

For each subspace  $W$  of dimension  $k \geq 2$  of  $V^*$ , consider the subspace of  $\text{End}(V^*)$  consisting of those endomorphisms whose restriction to  $W$  is a scalar endomorphism of  $W$ . The dimension of this space is  $1 + (r-k)^2 + k(r-k)$ . Varying  $W$  over the Grassmannian  $G(r,k)$  we get a vector bundle  $W(r,k)$  over  $G(r,k)$  and the dimension of the total space of this bundle is

$$1 + (r-k)^2 + k(r-k) + k(r-k) = 1 + r^2 - k^2.$$

We have a natural morphism  $\pi_k: W(r,k) \rightarrow \text{End}(V)$  which maps an endomorphism to its transpose. If  $S_k := \pi_k(W(r,k))$ , we have  $\dim S_k \leq (1 + r^2 - k^2)$  and  $\text{codim } S_k \geq k^2 - 1 \geq 3$ . Let  $S = \bigcup_{2 \leq k \leq r} S_k$  and  $U = \text{End}(V) - S$ . We have  $\text{codim } S \geq 3$  and  $S$  is saturated for the map  $\text{End}(V) \rightarrow \text{End}^0(V)$ . The fibres of  $q:q^{-1}(U) \rightarrow U$  are finite and  $q$  is proper. Hence  $q$  is a finite morphism.

**3. Sections of the (twisted) relative tangent bundle of a projective bundle**

Let  $E$  be a vector bundle of rank  $r \geq 2$  on a smooth irreducible projective variety  $X$  of dimension  $\geq 1$  over  $K$ . Let  $\pi:P(E) \rightarrow X$  be the associated projective bundle. We have the exact sequences on  $P(E)$ :

$$0 \rightarrow \Omega^1_\pi(1) \rightarrow \pi^*(E) \rightarrow \mathcal{O}_{P(E)}(1) \rightarrow 0$$

and

$$0 \rightarrow \mathcal{O}_{P(E)} \rightarrow \pi^*(E^*) \otimes \mathcal{O}(1) \rightarrow \Theta_\pi \rightarrow 0$$

where  $\Theta_\pi$  (resp.  $\Omega^1_\pi$ ) denotes the relative tangent (resp. cotangent) bundle along the fibres of  $\pi$ . Let  $\text{End}^0(E)$  denote the vector bundle  $\text{End}(E)/\mathcal{O}_X$ . We have an exact sequence of vector bundles on  $P(E)$ :

$$0 \rightarrow F \rightarrow \pi^*(\text{End}^0(E)) \rightarrow \Theta_\pi \rightarrow 0.$$

$$0 \rightarrow F \otimes \pi^*(M) \rightarrow \pi^*(\text{End}^0(E) \otimes M) \rightarrow \Theta_\pi \otimes \pi^*(M) \rightarrow 0$$

### PROPOSITION 3.1

Let  $p: F \otimes \pi^*(M) \rightarrow \text{End}^0(E) \otimes M$  be the canonical morphism (of total spaces of geometric vector bundles over  $\mathbf{P}(E)$  and  $X$  respectively). Then there exists an open subset  $\Omega$  of  $\text{End}^0(E) \otimes M$  whose complement is of codimension  $\geq 3$  such that the morphism  $p: p^{-1}(\Omega) \rightarrow \Omega$  is finite.

*Proof.* This follows from Proposition 2.1.

### PROPOSITION 3.2

There exists an ample line bundle  $M$  on  $X$  such that a generic section  $s$  of  $\Theta_\pi \otimes \pi^*(M)$  (i.e. for  $s$  in a non-empty open subset of  $H^0(\mathbf{P}(E), \Theta_\pi \otimes \pi^*(M))$ ) satisfies the following conditions:

- a) The scheme  $Z$  of zeros of  $s$  is smooth and irreducible.
- (b) The morphism  $\pi|_Z: Z \rightarrow X$  is surjective and separable.
- (c) There exists a closed subset  $S$  of  $X$  of codimension  $\geq 3$  such that the morphism

$$\pi: Z \setminus \pi^{-1}(S) \rightarrow X \setminus S$$

is finite.

*Proof.* Let  $\xi$  be an ample line bundle on  $X$ . Then the line bundle  $\pi^*(\xi^k) \otimes \mathcal{O}(1)$  is very ample on  $\mathbf{P}(E)$  for  $k \geq k_0$ . [4, II, Prop. 7.10, p. 161]. We may also assume that for  $k \geq k_0$ , the bundle  $\xi^k \otimes E^*$  is generated by its sections. Let  $M = \xi^{2k}$ . Since the sections of the very ample line bundle  $\pi^*(\xi)^k \otimes \mathcal{O}(1)$  generate its first order jet bundle and  $\pi^*(\xi^k \otimes E^*)$  is generated by its sections, we see that the sections of  $\pi^*(M \otimes E^*) \otimes \mathcal{O}(1)$  generate its first order jet bundle [7, Lemma 5]. Since  $\Theta_\pi \otimes \pi^*(M)$  is a quotient bundle of  $\pi^*(M \otimes E^*) \otimes \mathcal{O}(1)$ , the sections of  $\Theta_\pi \otimes \pi^*(M)$  generate its first order jet bundle. Now by [7, Theorem 1] the zero scheme  $Z$  of a generic section  $s$  of  $\Theta_\pi \otimes \pi^*(M)$  is smooth. We will prove in the next proposition (Proposition 4.1) that  $Z$  is irreducible.

Let  $x_0 \in X$  and  $S := \pi^{-1}(x_0)$  the fibre over  $x_0$ . Let  $W$  be the image of the homomorphism

$$H^0(\mathbf{P}(E), \Theta_\pi \otimes \pi^*(M)) \rightarrow H^0(S, \Theta_\pi \otimes \pi^*(M)|_S).$$

(We may even assume that  $W = H^0(S, \Theta_\pi \otimes \pi^*(M)|_S)$ , by choosing  $k$  large enough). Then the first order jets of elements of  $W$  generate the first order jet bundle of  $\Theta_\pi \otimes \pi^*(M)|_S$ ; hence, again by [7, Theorem 1]; for a generic element  $\sigma$  of  $W$ , the zero subscheme (of  $S$ ) defined by  $\sigma$  is smooth. Thus we see that for a generic section of  $\Theta_\pi \otimes \pi^*(M)$  the zero scheme  $Z$  is smooth and irreducible and  $Z$  intersects  $S$  transversally. It follows that there exists a point  $z_0 \in Z \cap S$  such that the differential of  $\pi|_Z$  at  $z_0$  is an isomorphism. This proves that  $\pi|_Z: Z \rightarrow X$  is surjective and (assuming  $Z$  to be irreducible) separable. (Observe that  $Z$  intersects every fibre  $\pi^{-1}(x)$ ,  $x \in X$ ,



line bundle.)

$$\begin{aligned}\text{Let } \Sigma &:= H^0(\mathbf{P}(E), \pi^*(M) \otimes \Theta_\pi) \\ &= H^0(X, M \otimes \pi_*(\Theta_\pi)) \\ &= H^0(X, M \otimes \text{End}^0(E)).\end{aligned}$$

Consider the morphisms

$$\begin{array}{c} \varphi: \Sigma \times X \rightarrow M \otimes \text{End}^0(E) \\ \downarrow p_X \\ X. \end{array}$$

where the evaluation map  $\varphi$  is a smooth morphism, being a surjection of vector bundles. Let  $\Omega$  be the open subset of  $M \otimes \text{End}^0(E)$  defined in Proposition 3.1 and  $N$  its complement. Then

$$\dim \varphi^{-1}(N) \leq \dim \Sigma + \dim X - 3.$$

By considering  $p_X: \varphi^{-1}(N) \rightarrow X$  we see that for a generic section  $s$  of  $M \otimes \text{End}^0(E)$  we have

$$\dim(\varphi^{-1}(N) \cap \{s \times X\}) \leq (\dim X) - 3.$$

Let  $S \subset X$  be defined to be  $p_X(\varphi^{-1}(N) \cap \{s \times X\})$ . Then  $\dim S \leq (\dim X) - 3$  and

$$\pi|_Z: Z \setminus \pi^{-1}(S) \rightarrow X \setminus S$$

is finite.

Thus for a generic section  $s$  of  $\Theta_\pi \otimes \pi^*(M)$  all the conditions a), b) and c) are satisfied.

#### 4. Koszul resolution of the zero scheme $Z$

*Proof of Theorem 4.2*

##### PROPOSITION 4.1

Let  $s$  be a section of  $\Theta_\pi \otimes \pi^*(M)$  over  $\mathbf{P}(E)$  such that the zero scheme  $Z$  of  $s$  is smooth. We then have

- $\pi_*(\mathcal{O}_Z \otimes \mathcal{O}_{\mathbf{P}(E)}(1)) \simeq E$  and  $R^i \pi_*(\mathcal{O}_Z(1)) = 0$  for  $i \geq 1$ .
- $\pi_*(\mathcal{O}_Z)$  has a filtration  
 $0 = F_0 \subset F_1 \subset F_2 \subset \dots \subset F_i \subset \dots \subset F_r = \pi_*(\mathcal{O}_Z)$   
such that  $F_i/F_{i-1} \simeq M^{-(i-1)} (= (M^*)^{\otimes (i-1)})$  for  $1 \leq i \leq r$  (In particular  $F_1 \simeq \mathcal{O}_Z$ ).
- $Z$  is irreducible (if  $\dim X \geq 1$  and  $M$  is ample).

*Proof.* Using our assumption on  $Z$ , we have a Koszul resolution for  $\mathcal{O}_Z$  on  $\mathbf{P}(E)$ : [1, Ch I, Lemma 4.2 and Ch. III, Propositions 4.10 and 4.11]

$$(A) \quad 0 \rightarrow \bigwedge^{r-1} (\Omega_\pi^1 \otimes \pi^*(M^*)) \rightarrow \dots \rightarrow \bigwedge^2 (\Omega_\pi^1 \otimes \pi^*(M^*)) \rightarrow \Omega_\pi^1 \otimes \pi^*(M^*) \rightarrow \mathcal{O}_{\mathbf{P}(E)} \rightarrow \mathcal{O}_Z \rightarrow 0$$

and also a resolution of  $\mathcal{O}_Z(1)$ :

$$(B) \quad 0 \rightarrow \Omega_{\pi}^{r-1} \otimes (\pi^*(M^*))^{r-1} \otimes \mathcal{O}(1) \rightarrow \cdots \rightarrow \Omega_{\pi}^1 \otimes \pi^*(M^*) \otimes \mathcal{O}(1) \rightarrow \mathcal{O}(1) \rightarrow \mathcal{O}(1)|_Z \rightarrow 0$$

Now we have, for the projective space  $\mathbf{P}$ , the Bott vanishing theorem:

$$H^i(\mathbf{P}, \Omega^p(1)) = 0 \text{ for } p \geq 1 \text{ and all } i \text{ [6, Théorème 1].}$$

Hence we have

$$R^i \pi_*(\Omega_{\pi}^p \otimes \pi^*((M^*)^{\otimes p}) \otimes \mathcal{O}(1)) = 0$$

for  $p \geq 1$  and all  $i$ . Splitting  $B$  into short exact sequences we deduce that

$$\pi_*(\mathcal{O}_Z(1)) = \pi_*(\mathcal{O}_{P(E)}(1)) = E$$

and  $R^i \pi_*(\mathcal{O}_Z(1)) = 0$  for  $i > 0$ .

For proving (b) we observe that  $R^q \pi_*(\Omega_{\pi}^p) = 0$  for  $p \neq q$  and

$$R^p \pi_*(\Omega_{\pi}^p) = \mathcal{O}_X \text{ for } 0 \leq p \leq (r-1) \text{ [6].}$$

Splitting (A) into short exact sequences we obtain b).

To prove c), since  $\dim X \geq 1$  and  $M$  is ample we have  $H^0(X, M^{-k}) = 0$  for  $k > 0$ . Using the filtration of  $\pi_*(\mathcal{O}_Z)$  given in b), we see that

$$H^0(Z, \mathcal{O}_Z) = H^0(X, \pi_*(\mathcal{O}_Z)) = H^0(X, \mathcal{O}_X) = K.$$

Since  $Z$  is smooth it follows that  $Z$  is irreducible.

**Theorem 4.2** *Let  $X$  be a smooth irreducible projective variety of dimension  $\geq 1$  over an algebraically closed field  $K$ . Let  $E$  be a vector bundle on  $X$ . Then there exist a smooth irreducible projective variety  $Z$  over  $K$ , a line bundle  $L$  on  $Z$  and a surjective separable morphism  $f: Z \rightarrow X$  having in addition the following properties:*

1) *there exists a closed subset  $S$  in  $X$  of codimension  $\geq 3$  such that the morphism*

$$f: Z \setminus f^{-1}(S) \rightarrow X \setminus S$$

*is finite.*

2) *we have  $f_*(L) \simeq E$  and  $R^i f_*(L) = 0$  for  $i > 0$ .*

*Proof.* We may assume that  $E$  is of rank  $\geq 2$ . Choose an ample line bundle  $M$  on  $X$  satisfying the conditions of Proposition 3.2. Let  $L$  be the restriction of  $\mathcal{O}_{P(E)}(1)$  to  $Z$  and  $f$  be the restriction of  $\pi: \mathbf{P}(E) \rightarrow X$  to  $Z$ . Then by Proposition 4.1(a) we have

$$f_*(L) \simeq E \text{ and } R^i f_*(L) = 0 \text{ for } i > 0.$$

## 5. The map $D$

Let  $f: Z \rightarrow X$  be a morphism and  $L$  a line bundle on  $Z$  such that  $f_*(L) = E$  is a vector bundle of rank  $r$  on  $X$ . The morphism  $f^*(f_*(L)) \rightarrow L$  gives rise to a morphism

$$f_*(\mathcal{O}_Z) \otimes E \rightarrow f_*(L) = E$$

which may be viewed as a morphism

$$D: f_*(\mathcal{O}_Z) \rightarrow E^* \otimes E.$$

( $D$  gives the canonical  $f_*(\mathcal{O}_Z)$ -module structure on  $f_*(L)$ ).

Suppose that  $f: Z \rightarrow X$  is a finite surjective morphism of smooth varieties. Then  $f$  is flat [1, Ch. V, Cor. 3.6]. We have

$$H^1(Z, \mathcal{O}_Z) = H^1(X, f_*(\mathcal{O}_Z)).$$

The homomorphism

$$D_*: H^1(Z, \mathcal{O}_Z) \rightarrow H^1(X, \text{End } E) \quad (5.1)$$

induced by  $D$  is the infinitesimal deformation map (at  $L$ ) for the variation of the direct image bundles as  $L$  deforms as a line bundle on  $X$  [2, Lemma 1.3.1].

Since  $f$  is finite, the map  $f^*(E) \rightarrow L$  is surjective and we have an exact sequence

$$0 \rightarrow N \rightarrow f^*(E) \rightarrow L \rightarrow 0$$

of vector bundles on  $Z$ . From this we get an exact sequence of vector bundles.

$$0 \rightarrow \mathcal{O}_Z \rightarrow f^*(E^*) \otimes L \rightarrow N^* \otimes L \rightarrow 0.$$

Since  $f$  is flat and finite,  $f_*(\mathcal{O}_Z)$  is a vector bundle on  $X$  of rank  $r$  and  $f_*(N^* \otimes L)$  is a vector bundle. So we have an exact sequence of vector bundles on  $X$ :

$$0 \rightarrow f_*(\mathcal{O}_Z) \rightarrow E^* \otimes E \rightarrow f_*(N^* \otimes L) \rightarrow 0.$$

Observe that  $f_*(\mathcal{O}_Z)/\mathcal{O}_X$  is a vector subbundle of rank  $(r-1)$  of the vector bundle  $\text{End}(E)/\mathcal{O}_X = \text{End}^0(E)$ . Thus we have

**Lemma 5.2** *Let  $f: Z \rightarrow X$  be a finite morphism of smooth varieties and  $L$  a line bundle on  $Z$ . If  $E = f_*(L)$  is a vector bundle of rank  $r$ , then the vector bundle  $\text{End}^0(E)$  contains a vector subbundle of rank  $(r-1)$ .*

Let us get back to the situation in §3 and §4.

### PROPOSITION 5.3

*Let  $s$  be a section of  $\Theta_\pi \otimes \pi^*(M)$  and  $Z$  the zero subscheme of  $s$  with the property that the canonical map  $E = \pi_*(\mathcal{O}_{\mathbf{P}(E)}(1)) \rightarrow f_*(\mathcal{O}_Z(1))$  is an isomorphism, where  $f = \pi|_Z$ . Suppose that  $T$  is a section of  $\text{End}(E) \otimes M$  such that the image of  $T$  in  $H^0(\mathbf{P}(E), \Theta_\pi \otimes \pi^*(M))$  is  $s$ . Then there is a homomorphism  $\mu: M^{-1} \rightarrow f_*(\mathcal{O}_Z)$  and a commutative diagram*

$$\begin{array}{ccc} f_*(\mathcal{O}_Z) & \xrightarrow{D} & \text{End } E \\ \mu \swarrow & & \nearrow T \\ & M^{-1} & \end{array}$$

where  $D$  is defined at the beginning of this section (§5).

$$\begin{array}{ccccccc}
 & & & \downarrow \tilde{T} & & & \\
 0 & \longrightarrow & \Omega_{\pi}^1 & \longrightarrow & \pi^*(E) & \xrightarrow{g} & \mathcal{O}(1) \longrightarrow 0
 \end{array}$$

where  $\tilde{T}$  is induced by  $T$ . The homomorphism  $g \circ \tilde{T} \circ i: \Omega_{\pi}^1 \otimes M^{-1} \rightarrow \mathcal{O}(1)$  gives the section  $s$  of  $\Omega_{\pi} \otimes \pi^*(M)$ . So  $\tilde{T}$  induces on  $Z$  a homomorphism  $\lambda: \mathcal{O}_Z(1) \otimes f^*(M^{-1}) \rightarrow \mathcal{O}_Z(1)$  and we have a commutative diagram

$$\begin{array}{ccc}
 \pi^*(E) \otimes \pi^*(M^{-1}) & \xrightarrow{q \otimes 1} & \mathcal{O}_Z(1) \otimes \pi^*(M^{-1}) \\
 \downarrow \tilde{T} & & \downarrow \lambda \\
 \pi^*(E) & \xrightarrow{q} & \mathcal{O}_Z(1)
 \end{array}$$

Considering  $\lambda$  as a section of  $\mathcal{O}_Z \otimes \pi^*(M)$  we obtain the section  $\pi_*(\lambda)$  of  $M \otimes \pi_*(\mathcal{O}_Z)$  which we view as a homomorphism  $\mu: M^{-1} \rightarrow f_*(\mathcal{O}_Z)$ . Since by assumption  $E \rightarrow \pi_*(\mathcal{O}_Z(1))$  is an isomorphism, it follows, from the above diagram, that  $T$  corresponds to  $\pi_*(\lambda): M^{-1} \otimes E \rightarrow E$ . But by the definitions of  $D$  and  $\mu$  we see that  $\pi_*(\lambda)$  corresponds to  $D \circ \mu$ . This means that  $T = D \circ \mu$ .

## 6. Vector bundles on curves

**Lemma 6.1.** *Let  $X$  be a smooth, projective irreducible curve over  $K$  and  $F$  a vector bundle of rank  $\geq 2$  on  $X$ . Let  $\xi$  be an ample line bundle on  $X$ . Then there exists an integer  $l_0$  such that for  $l \geq l_0$ ,  $\xi^{-l}$  is a subbundle of  $F$  and the induced homomorphism  $H^1(X, \xi^{-l}) \rightarrow H^1(X, F)$  is surjective.*

*Proof.* Let first  $F$  be of rank 2. Since for all large  $l$ ,  $\xi^l \otimes F^*$  contains a trivial line subbundle, we get an exact sequence

$$0 \rightarrow \xi^{-l} \rightarrow F \rightarrow \xi^l \otimes \det F \rightarrow 0.$$

Choose  $l$  large enough so that  $H^1(X, \xi^l \otimes \det F) = 0$ .

Now suppose that  $F$  is a vector bundle of rank  $\geq 3$ . Then we can find a filtration of  $F$  by subbundles

$$0 \subset F_1 \subset F_2 \subset \dots \subset F_i \subset \dots \subset F_{r-1} = F$$

such that  $\text{rank}(F_i) = i + 1$  (in particular  $\text{rank } F_1 = 2$ ) and such that  $H^1(X, F_i/F_{i-1}) = 0$ , for  $i \geq 2$ . Now choose a line subbundle  $\xi^{-l}$  of  $F_1$  with  $H^1(F_1/\xi^{-l}) = 0$ . We see easily, by induction on  $i$ , that  $H^1(X, F/\xi^{-l}) = 0$ .

**Remark 6.2** Note that  $H^1(X, \xi^{-l}) \rightarrow H^1(X, F)$  is surjective if and only if  $H^1(F/\xi^{-l}) = 0$ , as  $H^2(X, \xi^{-l}) = 0$ .

# COROLLARY 6.3

Let  $F$  be a vector bundle on  $X$ . Then there exists an integer  $l_0$  such that, for  $l \geq l_0$ , for a generic section  $\sigma$  of  $\xi^l \otimes F$  the map  $H^1(X, \xi^{-l}) \rightarrow H^1(X, F)$  induced by  $\sigma$  is surjective.

**Theorem 6.4** Let  $X$  be a smooth projective irreducible curve over an algebraically closed field  $K$  and let  $E$  be a vector bundle on  $X$ . Then there exist a smooth projective irreducible curve  $Z$  over  $K$ , a line bundle  $L$  on  $Z$  and a finite surjective separable morphism  $f: Z \rightarrow X$  such that

$$f_*(L) \simeq E$$

and the homomorphism (defined in 5.1)

$$H^1(Z, \mathcal{O}_Z) \rightarrow H^1(X, \text{End } E)$$

is surjective.

*Proof.* Choose an ample line bundle  $M$  as in the proof of Theorem 4.2. We may also choose  $M$  to have the further properties:

a)  $H^1(X, M) = 0$

and

b) a generic section  $\sigma$  of  $H^0(X, \text{End } E \otimes M)$  verifies the condition that the homomorphism

$$H^1(X, M^{-1}) \rightarrow H^1(X, \text{End } E)$$

is surjective (use Corollary 6.3).

By condition a) the map

$$H^0(X, \text{End } E \otimes M) \rightarrow H^0(X, \text{End}^0 E \otimes M)$$

is surjective.

Now a generic section  $s$  of  $H^0(\mathbf{P}(E), \Theta_\pi \otimes \pi^*(M)) = H^0(X, \text{End}^0(E) \otimes M)$  is the image of a section  $T$  of  $\text{End}(E) \otimes M$  with the property that the homomorphism

$$H^1(X, M^{-1}) \rightarrow H^1(X, \text{End } E)$$

induced by  $T$  is surjective and satisfies conditions a), b) and c) of Proposition 3.2.

Choose  $Z, L, f$  as in the proof of Theorem 4.2. Then  $f_*(L) = E$ .

To prove 2), observe that the factorisation, given in Proposition 5.3,

$$\begin{array}{ccc} f_*(\mathcal{O}_Z) & \xrightarrow{D} & \text{End } E \\ \mu \swarrow & & \nearrow T \\ & M^{-1} & \end{array}$$

induces a commutative diagram:

$$\begin{array}{ccccc} H^1(Z, \mathcal{O}_Z) & = & H^1(X, f_*(\mathcal{O}_Z)) & \xrightarrow{D_*} & H^1(X, \text{End } E) \\ \mu_* \swarrow & & & \nearrow T_* & \\ & H^1(X, M^{-1}) & & & \end{array}$$

**Remark 6.5** If  $E$  is a stable bundle on  $X$  and if  $Z, f$  and  $L$  are chosen as in Theorem 6.4, we see that  $f_*$  gives a *dominant* separable rational morphism from an appropriate component of  $\text{Pic}(Z)$  into the moduli space of vector bundles on  $X$  of rank  $r = rk E$  and degree  $d = \text{degree } E$  (compare [3]). We thus obtain 'most' stable bundles of a given rank and degree as direct images of line bundles on a *fixed* covering  $Z$  of  $X$ .

## 7. The example

We now give an example of a rank 2 vector bundle on the projective space  $\mathbf{P}_3(\mathbf{C})$  which cannot be obtained as the direct image of a line bundle by a *finite* morphism  $f: Z \rightarrow \mathbf{P}_3(\mathbf{C})$ , with  $Z$  smooth.

Let  $E$  be a stable vector bundle of rank 2 on  $\mathbf{P}_3(\mathbf{C})$  with  $c_1(E) = 0$  and  $c_2(E) > 0$ . If  $E$  were the direct image of a line bundle by  $f: Z \rightarrow X$ , with  $Z$  smooth and  $f$  finite, the bundle  $\text{End}^0(E)$  would contain a line *subbundle*  $L$  by Lemma 5.2. If  $\xi = L^{-1}$ , we would have

$$c_3(\xi \otimes \text{End}^0 E) = 0. \text{ We have}$$

$$c_3(\xi \otimes \text{End}^0 E) = 4c_1(\xi)c_2(E) + c_1(\xi)^3.$$

But the bundle  $L$  and hence  $\xi$ , is non-trivial, since  $h^0(\mathbf{P}_3, \text{End}^0 E) = 0$ ,  $E$  being stable. So  $c_1(\xi) \neq 0$  and we would have

$$c_1(\xi)(4c_2(E) + c_1(\xi)^2) = 0,$$

a contradiction.

## Acknowledgements

The first author (AH) carried out this work in the framework of the Vector Bundle group of Europroj. The second author (MSN) would like to thank CNRS and Université de Nice-Sophia Antipolis for hospitality when part of this work was done.

## References

- [1] Altman A and Kleiman S, *Introduction to Grothendieck duality theory* LNM 146 (1970) Springer
- [2] Beauville A, Fibrés de rang 2 sur une courbe, fibré déterminant et fonction theta, *Bull. Soc. Math. France*, **116** (1988) 431–418
- [3] Beauville A, Narasimhan M S and Ramanan S, Spectral curves and the generalised theta divisor, *J. Reine Angew. Math.* **398** (1989) 169–179
- [4] Hartshorne R, *Algebraic Geometry* (Springer-Verlag) (1977)
- [5] Schwarzenberger R L E, Vector bundles on the projective plane, *Proc. London Math. Soc.* (3) **11** (1961) 623–640
- [6] Verdier J L, Le théorème de Le Potier, Séminaire de géométrie analytique, *Astérisque* **17** (1974)
- [7] Walter C H, Transversality theorems in general characteristic and arithmetically Buchsbaum schemes *Int. J. Math.* (to appear)

# Finite arithmetic subgroups of $GL_n$ , III

YOSHIYUKI KITAOKA

Department of Mathematics, School of Science, Nagoya University, Japan

Dedicated to the memory of Professor K G Ramanathan

**Abstract.** Let  $G$  be an algebraic group in  $GL_n(\mathbb{C})$  defined over  $\mathbb{Q}$ , and  $K$  an algebraic number field with the maximal order  $O_K$ . If the group  $G(O_K)$  of rational points of  $G$  in  $M_n(O_K)$  is a finite group and if it satisfies a certain condition, which is satisfied, for example, when  $K$  is a nilpotent extension of  $\mathbb{Q}$  and 2 is unramified, then  $G(O_K)$  is generated by roots of unity in  $K$  and  $G(\mathbb{Z})$ .

**Keywords.** Algebraic group; algebraic number field; quadratic form; finite arithmetic subgroup.

Let  $K$  be a Galois extension of the rational number field  $\mathbb{Q}$  with Galois group  $G(K/\mathbb{Q})$  and  $O_K$  the maximal order of  $K$ . Let  $G$  be a finite subgroup in  $GL_n(O_K)$  such that  $G$  is stable under the action of  $G(K/\mathbb{Q})$ , that is  $u(g) := (u(g_{ij})) \in G$  for every  $g := (g_{ij}) \in G$  and  $u \in G(K/\mathbb{Q})$ . Our problem is whether  $G$  is of  $A$ -type in the sense of [2], that is whether there is an element  $h \in GL_n(\mathbb{Z})$  such that

$$\{hgh^{-1} | g \in G\} \subset \{\text{diag}(\varepsilon_1 A_1, \dots, \varepsilon_m A_m) | A_i \in GL_{n_i}(\mathbb{Z}), \varepsilon_i: \text{root of unity}\}.$$

Here  $\text{diag}(\varepsilon_1 A_1, \dots, \varepsilon_m A_m)$  denotes the matrix in which  $\varepsilon_1 A_1, \dots, \varepsilon_m A_m$  are diagonally arranged. If, hence  $\{\pm 1\}$  are the only roots of unity in  $K$ , then  $G$  being of  $A$ -type means  $G \subset GL_n(\mathbb{Z})$ .

The lattice-theoretic meaning is: Let  $L$  be a free module of rank  $n$  over  $\mathbb{Z}$ , and let  $\mathcal{G}$  be a linear algebraic group in  $GL_n$  defined over  $\mathbb{Q}$ . Suppose that  $K$  is a Galois extension of  $\mathbb{Q}$  and that  $\mathcal{G}(O_K)$  is a finite group, which is canonically regarded as a subgroup of automorphisms of  $O_K L$ . Then  $\mathcal{G}(O_K)$  being of  $A$ -type implies that there exists a direct sum decomposition  $L = \bigoplus_{i=1}^k L_i$  so that every  $\sigma \in \mathcal{G}(O_K)$  and for roots of unity  $\varepsilon_i \in K$  dependent on  $\sigma$ , we have

$$\varepsilon_i \sigma(L_i) = L_i \text{ for } i = 1, \dots, k.$$

We know that the above question is affirmative if either  $K$  is totally real and  $G(K/\mathbb{Q})$  is nilpotent or  $G(K/\mathbb{Q})$  is abelian ([1], [2]). The aim of this paper is to show that this is affirmative if  $G(K/\mathbb{Q})$  is nilpotent and the complex conjugation induces an element of the center of  $G(K/\mathbb{Q})$ . In § 2, we give miscellaneous results.

## 1. Main result

*Lemma 1.* Let  $K/\mathbb{Q}$  be a Galois extension with Galois group  $\Gamma := G(K/\mathbb{Q})$ . Denote by  $W$  the set of all roots of unity in  $K$ . We fix an element  $a_\sigma \in W$  for  $\sigma \in \Gamma$ . If  $a_{\mu\sigma} = a_\mu a_\sigma$

holds for every  $\sigma, \mu \in \Gamma$ , then there exists an element  $a \in K$  such that  $a_\mu = \mu(a^{-1})a$  for  $\mu \in \Gamma$  and  $a^w \in Q^*$ , where  $w$  is the cardinality of  $W$ .

*Proof.* We can take a non-zero element  $u \in K$  such that  $a := \sum_{\sigma \in \Gamma} a_\sigma \sigma(u) \neq 0$ . For  $\mu \in \Gamma$ , we have  $\mu(a) = \sum_{\sigma \in \Gamma} \mu(a_\sigma) \mu \sigma(u) = \sum_{\sigma \in \Gamma} a_\mu^{-1} a_{\mu\sigma} \mu \sigma(u) = a_\mu^{-1} a$ , and hence  $a_\mu = \mu(a^{-1})a$ .  $a_\mu \in W$  implies  $a_\mu^w = 1$  and therefore  $\mu(a^w) = a^w$ . This yields  $a^w \in Q^*$ . ■

**Lemma 2.** Let  $p$  be a prime number and let  $A \in GL_n(\mathbf{Z})$  be of finite order and suppose that  $A \equiv 1_n \pmod{p}$ . If  $p \neq 2$ , then we have  $A = 1_n$ . If  $p = 2$ , then  $A = T \begin{pmatrix} 1 & \\ & -1 \end{pmatrix} T^{-1}$  for some  $T \in GL_n(\mathbf{Z})$ .

*Proof.* This is due to Minkowski. But we give the proof for the convenience. For  $B \in GL_n(\mathbf{Z})$  with  $B \equiv 1_n \pmod{p}$ , we write  $B = 1_n + p^r C$ , where  $C$  is integral and  $C \not\equiv 0 \pmod{p}$ . Then for a natural number  $h$ , the following is clear:

$$B^h = 1_n + hp^r C + \sum_{k=2}^h \binom{h}{k} (p^r C)^k. \quad (1)$$

Suppose that the order of  $A$  is not a power of  $p$ . Then some power  $B$  of  $A$  is of order  $h$ , where  $h$  is a prime different from  $p$ . Eqn. (1) implies  $1_n \equiv 1_n + hp^r C \pmod{p^{2r}}$ , which contradicts  $C \not\equiv 0 \pmod{p}$ . Thus the order of  $A$  is a power of  $p$ .

Suppose  $p \neq 2$  and  $A \neq 1_n$ . Let  $B$  be a power of  $A$  whose order is  $p$ . Applying (1) with  $h = p$ ,  $1_n \equiv 1_n + p^{r+1} C \pmod{p^{2r+1}}$  follows from  $r \geq 1$  and  $\binom{p}{k} \equiv 0 \pmod{p}$  for  $k \neq 0, p$ . This contradicts  $C \not\equiv 0 \pmod{p}$ , and so we have  $A = 1_n$  if  $p \neq 2$ .

Let us consider the case of  $p = 2$ .

Suppose that the order of  $A$  is 1 or 2. Let  $L$  be a  $\mathbf{Z}$ -module with basis  $\{e_1, \dots, e_n\}$  and define a linear mapping  $u$  by  $(u(e_1), \dots, u(e_n)) = (e_1, \dots, e_n)A$ . Then  $A \equiv 1_n \pmod{2}$  and  $A^2 = 1_n$  imply  $u(x) \equiv x \pmod{2L}$  and  $u^2 = \text{id}$ . Hence  $x = (x + u(x))/2 + (x - u(x))/2$  implies  $L = \{x \in L \mid u(x) = x\} \oplus \{x \in L \mid u(x) = -x\}$ . Therefore this completes the proof if  $A^2 = 1_n$ .

Lastly, supposing that the order of  $A = 2^k$  ( $k \geq 2$ ), we will show the contradiction. Applying (1) for  $B = A$ , we have  $A^2 \equiv 1_n \pmod{4}$ . Hence in the expression of  $B := A^{2^{k-1}} = 1_n + 2^r C$ , we have  $r \geq 2$ . Then we have  $1_n = B^2 = 1_n + 2^{r+1} C + 2^{2r} C^2$ , which yields the contradiction  $C \equiv 0 \pmod{2^{r-1}}$ . Thus we have completed the proof. ■

**Lemma 3.** Let  $K/Q$  be a Galois extension with Galois group  $\Gamma := G(K/Q)$  and  $O_K$  the maximal order of  $K$ . Let  $G$  be a finite subgroup of  $GL_n(O_K)$  which is stable under the action of  $\Gamma$ . Let  $\Gamma'$  be the commutator subgroup of  $\Gamma$  and  $K'$  be the maximal abelian subfield of  $K$  corresponding to  $\Gamma'$ . Suppose the following conditions:

(I) if  $F$  is a proper subfield of  $K$  and  $F/Q$  is a Galois extension, then  $G \cap GL_n(F) \subset GL_n(K')$ .

(II) At least two rational primes ramify in  $K$ .

Then  $G$  is of  $A$ -type.

*Proof.* First we give a remark on elements in  $G' := G \cap GL_n(K')$ . Putting  $P := \sum_{g \in G'} g \bar{g}$ ,  $\bar{P} = P$  is a positive definite Hermitian matrix. Since  $K'$  is an abelian extension and  $\mu(G') = G'$  for  $\mu \in G(K'/Q)$ , we have  $\mu(P) = \sum_{g \in G'} \mu(g) \overline{\mu(g)} = P$  for  $\mu \in G(K'/Q)$ . Hence  $P$  is an integral positive definite symmetric matrix. Put  $L := \mathbf{Z}^n$  (column vectors) and



for  $x, y \in O_K, L = (O_K)$  we introduce an inner product by  $(x, y) := 'xP\bar{y}$ . Note that  $x, y \in L$  implies  $(x, y) \in \mathbb{Z}$ . Let  $L = \perp_{i=1}^m L_i$  be an orthogonal decomposition of  $L$  to indecomposable submodules. Then  $O_K \cdot L_i$  is also indecomposable with respect to the inner product by Lemma on p. 142 in [2]. Every  $g \in G'$  satisfies  $'gP\bar{g} = P$  and hence  $x \mapsto gx$  induces an isometry  $\sigma$  of  $O_K \cdot L$ . From theorems on p. 140 and p. 141 it follows that there exist roots of unity  $\varepsilon_i$  in  $K'$  such that  $\varepsilon_i \sigma(L_i) = L_i$ .

Take any element  $g \in G$ . We will show  $g \in GL_n(K')$ .

Let  $q$  be a rational prime which ramifies in  $K$ . We take a rational prime  $p$  different from  $q$  which also ramifies in  $K$ . Let  $\tilde{p}, \tilde{q}$  be prime ideals of  $K$  on  $p, q$  respectively. First, let us show that for every element  $\mu$  in the inertia group  $T(\tilde{q})$ , there is an integral matrix  $T \in GL_n(\mathbb{Z})$  dependent only on the above decomposition  $L = \perp L_i$  of  $L$  such that

$$g_1 := \mu(g)g^{-1} = TD_\mu T^{-1}$$

for a diagonal matrix  $D_\mu$  whose diagonal elements are roots of unity in  $K'$ .  $g_1 \equiv 1_n \pmod{\tilde{q}}$  follows from the definition and hence  $\sigma(g_1) = g_1$  for  $\sigma \in T(\tilde{p})$  by Lemma 7.5.2 in [3]. Considering other prime ideals lying on  $p, g_1$  is fixed by the subgroup  $\Gamma_1$  generated by inertia subgroups of prime ideals lying on  $p$ . Since  $\Gamma_1$  is normal in  $\Gamma$  and  $\Gamma_1 \neq \{1\}$ , the condition (I) yields  $g_1 \in GL_n(K')$ . By the remark at the beginning of the proof, there exist roots of unity  $\varepsilon_i$  such that  $\varepsilon_i \eta(L_i) = L_i$ , where  $\eta(x) := g_1 x$ . Put  $\mu := \varepsilon_i \eta$ . Then  $\mu(L_i) = L_i$  implies  $\eta(O_K \cdot L_i) = O_K \cdot L_i$  and then  $g_1 \equiv 1_n \pmod{\tilde{q}}$  implies  $\eta(x) \equiv x \pmod{\tilde{q}L_i}$  for  $x \in L_i$  and so  $\mu(x) \equiv \varepsilon_i x \pmod{\tilde{q}L_i}$  for  $x \in L_i$ . Comparing the coordinates,  $\varepsilon_i$  is congruent to some rational integer  $a$  modulo  $\tilde{q}$ . Since  $\varepsilon_i$  is a unit,  $a$  and  $q$  are relatively prime, and there is a rational integer  $b$  such that  $ab \equiv 1 \pmod{q}$ . Hence  $\eta(x) \equiv x \pmod{\tilde{q}L}$  for  $x \in L$  implies  $\mu(x) \equiv ab\mu(x) \equiv ab\varepsilon_i x \equiv a^2bx \equiv ax \pmod{\tilde{q}L_i}$  for  $x \in L_i$ , and hence  $\mu(x) \equiv ax \pmod{qL_i}$  for  $x \in L_i$ . Let  $\{f_1, \dots, f_m\}$  be a basis of  $L_i$  and define a matrix  $B \in GL_m(\mathbb{Z})$  by  $(\mu(f_1), \dots, \mu(f_m)) = (f_1, \dots, f_m)B$ . Then  $B \equiv a1_m \pmod{q}$  and the order  $k$  of  $\mu$  and hence  $B$  is finite. Put  $S = \sum_{r \pmod{k}} 'B^r B^r$ , and take an integer  $s$  such that  $S_0 := q^{-s}S$  is integral and is not congruent to 0 mod  $q$ . If  $q$  is odd, then there is a  $\mathbb{Z}$ -vector  $x$  such that  $'xS_0x \not\equiv 0 \pmod{q}$ . Hence we have  $'xS_0x = (Bx)S_0(Bx) \equiv a^2'xS_0x \pmod{q}$  and hence  $a^2 \equiv 1 \pmod{q}$ , which yields  $a \equiv \pm 1 \pmod{q}$ . This is true even for  $q = 2$ , since  $(a, q) = 1$ . Thus we have  $\mu(x) \equiv \pm x \pmod{qL_i}$  for  $x \in L_i$ . Since  $\mu = \varepsilon_i \eta$  is of finite order,  $\mu = \pm id$  on  $L_i$  follows from Lemma 2 and the indecomposability of  $L_i$ . If  $\mu = -id$ , then taking  $-\varepsilon_i$  instead of  $\varepsilon_i$ , we may assume that  $\varepsilon_i \eta$  is  $id$  on  $L_i$ . Taking the union of bases of  $L_i$ 's as a basis of  $L$ , and denoting the base change matrix by  $T \in GL_n(\mathbb{Z})$ , we can conclude that  $g_1 = \mu(g)g^{-1} = TD_\mu T^{-1}$  for a diagonal matrix whose diagonal entries are roots of unity in  $K$ . Thus (1) has been proved.

If  $\mu_i \in G(K/Q)$  ( $i = 1, 2$ ) satisfies  $\mu_i(g)g^{-1} = TD_{\mu_i} T^{-1}$ , then we have  $\mu_1 \mu_2(g)g^{-1} = \mu_1(TD_{\mu_2} T^{-1}g)g^{-1} = T\mu_1(D_{\mu_2})T^{-1}(TD_{\mu_1} T^{-1}) = T\mu_1(D_{\mu_2})D_{\mu_1} T^{-1}$  and  $\mu_1(D_{\mu_2})D_{\mu_1}$  is also a diagonal matrix whose diagonal elements are roots of unity in  $K$ . Noting that  $G(K/Q)$  is generated by inertia subgroups of all ramified prime ideals, we have  $\mu(g)g^{-1} = TD_\mu T^{-1}$  for  $\mu \in G(K/Q)$ , where  $D_\mu$  is a diagonal matrix whose diagonal entries are roots of unity in  $K$ . Since  $D_{\mu_1 \mu_2} = \mu_1(D_{\mu_2})D_{\mu_1}$  for  $\mu_1, \mu_2 \in G(K/Q)$ , Lemma 1 implies the existence of diagonal matrix  $D = \text{diag}(d_1, \dots, d_n) \in M_n(K)$  such that  $D_\mu = \mu(D^{-1})D$  and  $d_i^w \in \mathbb{Q}^\times$ , where  $w$  is the order of the group of roots of unity in  $K$ . Hence  $\mu(g)g^{-1} = TD_\mu T^{-1} = T\mu(D^{-1})DT^{-1}$  yields  $\mu(DT^{-1}g) = DT^{-1}g$  for  $\mu \in G(K/Q)$ . This means  $h := DT^{-1}g \in M_n(\mathbb{Q})$ . Taking a rational diagonal matrix  $h'$  so that the numbers on any row of  $h'h$  are relatively prime integers, then  $T^{-1}g = (h'D)^{-1}h'h \in M_n(O_K)$  implies that the diagonal matrix  $\tilde{D} := (h'D)^{-1}$  is in  $M_n(O_K)$ . Moreover

the fact that  $\det(T^{-1}g) = (\det(\tilde{D})\det(h'h))$  is in  $O_K^\times$  yields that diagonal entries  $\tilde{d}_i$  of  $\tilde{D}$  are also in  $O_K^\times$  by  $h'h \in M_n(\mathbb{Z})$ . Thus we have  $\tilde{d}_i \in O_K^\times$  and  $\tilde{d}_i^w \in \mathbb{Q}^\times$  by  $d_i^w \in \mathbb{Q}^\times$  and hence  $\tilde{d}_i$  is a root of unity. Thus we have proved that  $T^{-1}g = \tilde{D}h'h = (h'D)^{-1}h'h$  and hence  $g$  is in  $GL_n(K')$ . Thus  $G$  is in  $GL_n(O_K)$  and hence it is of  $A$ -type by Theorem on p. 141 in [2]. ■

**Theorem 1.** Let  $K$  be an algebraic number field such that (i)  $K/\mathbb{Q}$  is a nilpotent extension, and (ii) if  $H$  is a subfield of  $K$  such that  $H$  is a Galois extension over  $\mathbb{Q}$  and 2 is the only rational prime that ramifies in  $H$ , then the complex conjugation induces an element of the center of  $G(H/\mathbb{Q})$ . Let  $G$  be a finite subgroup of  $GL_n(O_K)$  which is stable under the action of  $G(K/\mathbb{Q})$ . Then  $G$  is of  $A$ -type.

*Proof.* We use induction on  $[K:\mathbb{Q}]$ . When  $[K:\mathbb{Q}] = 1$ , we have nothing to do. We note that for a proper subfield  $F$  of  $K$  which is a Galois extension of  $\mathbb{Q}$ , the conditions (i) and (ii) are satisfied. Suppose that there are at least two rational primes which ramify in  $K$ . Lemma 3 completes the proof, since the condition (I) in Lemma 3 is satisfied by the induction assumption. Hence we may suppose that there is only one rational prime  $p$  that ramifies in  $K$ . To complete the proof, we have only to claim that  $K$  is abelian by induction on  $[K:\mathbb{Q}]$ , assuming (i) and (ii), since the theorem is proved for every abelian field  $K$ .

Let  $Z$  be the center of  $G(K/\mathbb{Q})$ . Then we have  $Z \neq \{1\}$  and  $G(K/\mathbb{Q})/Z$  is nilpotent, and the subfield of  $K$  corresponding to  $Z$  satisfies the conditions (i) and (ii). By the induction assumption,  $G(K/\mathbb{Q})/Z$  is abelian and the complex conjugation is trivial by (ii) if  $p = 2$ . Thus  $G(K/\mathbb{Q})/Z$  is cyclic and hence  $G(K/\mathbb{Q})$  is abelian. ■

## 2. Miscellaneous results

**Lemma 4.** Let  $n$  be a natural number. Then there exists a finite set  $S_n$  of algebraic numbers with  $S_n \cap \mathbb{Q} = \phi$  which satisfies the following:

Let  $K$  be a Galois extension of degree  $n$  of  $\mathbb{Q}$  and suppose that  $S_n \cap K = \phi$  and the complex conjugation is in the center of  $G(K/\mathbb{Q})$ . Then the maximal order  $O_K$  is a positive lattice of  $E$ -type with respect to the quadratic form  $\text{tr}_{K/\mathbb{Q}}|x|^2$  in the sense of [3].

*Proof.* Denote by  $S_n$  the set of non-rational algebraic integers  $x$  satisfying that  $[\mathbb{Q}(x):\mathbb{Q}] \leq n$  and the square of the absolute value of every conjugate of  $x$  over  $\mathbb{Q}$  is less than  $(4/3)^{n-2} + 1/4$ . Then  $S_n$  is a finite set of algebraic integers and  $S_n \cap \mathbb{Q} = \phi$ . Let  $K$  be the field in the statement of Lemma. Then  $\text{tr}_{K/\mathbb{Q}}|x|^2 \geq 0$  for  $x \in O_K$  and the equality occurs if and only if  $x = 0$ . For  $x \in O_K$  ( $x \neq 0$ ), we have  $\text{tr}_{K/\mathbb{Q}}|x|^2 \geq n(\prod_{\sigma \in G(K/\mathbb{Q})} |\sigma(x)|^2)^{1/n} = n|N_{K/\mathbb{Q}}(x)|^{2/n} \geq n$  and  $\text{tr}_{K/\mathbb{Q}}|1|^2 = n$  is clear. If  $x \in O_K$  is not in  $\mathbb{Z}$ , then  $S_n \cap K = \phi$  implies that the absolute value of some conjugate of  $x^2$  is larger than  $(4/3)^{n-2} + 1/4$  and hence  $\text{tr}_{K/\mathbb{Q}}|x|^2 > (4/3)^{n-2} + 1/4$ . Let  $v_1 = 1, v_2, \dots, v_n$  be a basis of  $O_K$  over  $\mathbb{Z}$  such that the matrix  $(\text{tr}_{K/\mathbb{Q}} v_i v_j)$  is in the Siegel domain  $S_{4/3, 1/2}$ , that is  $(\text{tr}_{K/\mathbb{Q}} v_i v_j) = A[N]$ , where  $A = \text{diag}(a_1, \dots, a_n)$  with  $a_i/a_{i+1} \leq 4/3$  and  $N = (n_{ij})$  satisfying  $n_{ij} = 0$  if  $i > j$ ,  $= 1$  if  $i = j$  and  $|n_{ij}| \leq 1/2$  if  $i < j$ . Then we have

$$\begin{pmatrix} 1 & \overline{\text{tr}} v_2 \\ \overline{\text{tr}} v_2 & \overline{\text{tr}} |v_2|^2 \end{pmatrix} = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix} \begin{bmatrix} 1 & n_{12} \\ 0 & 1 \end{bmatrix}$$

and hence  $a_1 = 1, \text{tr } v_2 = a_1 n_{12} = n_{12}, \text{tr} |v_2|^2 = n_{12}^2 + a_2 \leq 1/4 + a_2$ .  $v_2 \notin \mathbf{Q}$  implies  $\text{tr}_{K/\mathbf{Q}} |v_2|^2 > (4/3)^{n-2} + 1/4$  and hence  $a_2 > (4/3)^{n-2}$ . Thus  $O_K$  is of  $E$ -type by Exercise 4 in §4 of Chapter 7 in [3]. ■

**Remark.** By Theorem 7.1.1 in [3], we can assume  $S_n = \phi$  if  $n \leq 43$ .

**Theorem 2.** Let  $n$  be a natural number. Then there exists a finite set  $S_n$  of non-rational algebraic integers which satisfies the following:

Let  $K$  be a Galois extension of degree  $n$  over  $\mathbf{Q}$  with  $S_n \cap K = \phi$  and assume that the complex conjugation is in the center of  $G(K/\mathbf{Q})$ . If  $G$  is a finite group of  $GL_n(O_K)$  which is stable under the action of  $G(K/\mathbf{Q})$ , then  $G$  is of  $A$ -type.

**Proof.** Let  $S_n$  be the set given in Lemma 4. Then  $O_K$  is a positive lattice of  $E$ -type by Lemma 4 and hence Theorem on p. 141 in [2] completes the proof. ■

**Remark.** The set  $S_n$  constructed in Lemma 1 contains  $r$ -th roots of unity for  $r \leq n$ . Therefore in Theorem 2 the conclusion " $G$  is of  $A$ -type" is replaced by " $G \subset GL_n(\mathbf{Z})$ ".

**Lemma 5.** Let  $K$  be a Galois extension of  $\mathbf{Q}$ . Let  $G$  be a subgroup of  $GL_n(O_K)$  stable under the action of  $G(K/\mathbf{Q})$ . For a prime ideal  $I$  of  $K$ , we put  $V_t(I, K/\mathbf{Q}) := \{u \in G(K/\mathbf{Q}) | u(x) \equiv x \pmod{I^{t+1}} \text{ for every } x \in O_K\}$  and  $G(I^i) := \{g \in G | g \equiv 1_n \pmod{I^i}\}$ . Suppose  $G(I^r) \neq \{1_n\}$  and  $G(I^{r+1}) = \{1_n\}$  for a natural number  $r \geq 1$ . If  $t, m$  are integers satisfying  $t \geq 0, m \geq 1$  and  $t + m \geq r + 1$ , then  $V_t(I, K/\mathbf{Q})$  acts trivially on  $G(I^m)$ , and  $G(I^s)$  is abelian if  $2s \geq r + 1$ .

**Proof.** Take an integer  $\pi \in O_K$  so that  $\pi I^{-1}$  is an integral ideal relatively prime to  $I$ . Let  $a, b$  be non-negative rational integers. We claim

$$u(\pi)^a \equiv \pi^a \pmod{I^{a+b}} \text{ if } u \in V_b(I, K/\mathbf{Q}).$$

When  $a = 1$ , this is contained in the definition. Suppose  $u(\pi)^a \equiv \pi^a \pmod{I^{a+b}}$  for  $a \geq 1$ . Write  $u(\pi) = \pi + x, u(\pi)^a = \pi^a + y$  where  $x \in I^{b+1}, y \in I^{a+b}$ . Then  $u(\pi)^{a+1} = (\pi + x)(\pi^a + y) = \pi^{a+1} + x\pi^a + y\pi + xy \equiv \pi^{a+1} \pmod{I^{a+b+1}}$  is clear. Thus we have completed the proof of the above claim.

Let  $t, m, r$  be rational integers such that  $t \geq 0, m \geq 1$  and  $t + m \geq r + 1$ . For  $g \in G(I^m)$ , we write  $g = 1_n + \pi^m A$  for  $A \in M_n(O_I)$ , where  $O_I$  denotes the  $I$ -adic completion of  $O_K$ . If  $u \in V_t(I, K/\mathbf{Q})$ , then we have

$$\begin{aligned} u(g) &= 1_n + u(\pi)^m u(A) \\ &= 1_n + (\pi^m \pmod{I^{m+t}})(A \pmod{I^{t+1}}) \\ &\equiv 1_n + \pi^m A \pmod{I^{t+m}} \\ &\equiv g \pmod{I^{t+m}}, \end{aligned}$$

which implies  $u(g)g^{-1} \in G(I^{t+m}) \subset G(I^{r+1}) = \{1_n\}$  and hence  $u(g) = g$ . Thus  $V_t(I, K/\mathbf{Q})$  acts trivially on  $G(I^m)$ .

$h = 1_n + \pi^s B(A, B \in M_n(O_I))$ , we have  $gh - hg = \pi^{2s}(AB - BA) \equiv 0 \pmod{I^{r+1}}$ . Hence  $ghg^{-1}h^{-1} \equiv 1_n \pmod{I^{r+1}}$  which means  $ghg^{-1}h^{-1} \in G(I^{r+1}) = \{1_n\}$ . Thus  $G(I^s)$  is abelian. ■

**Theorem 3.** Let  $K$  be a totally real Galois extension of  $\mathbb{Q}$  and suppose that there is a rational prime  $p$  which is totally ramified in  $K$ . If  $G$  is a finite subgroup of  $GL_n(O_K)$  stable under the action of  $G(K/\mathbb{Q})$ , then  $G$  is in  $GL_n(\mathbb{Z})$ .

*Proof.* First we note that the inertia subgroup  $V_0(I, K/\mathbb{Q})$  for the prime ideal  $I$  of  $K$  over  $p$  is equal to  $G(K/\mathbb{Q})$ , and  $N_{K/\mathbb{Q}}(I) = p$  is clear. If  $p = 2$ , then the order of the inertia group is a power of 2 and hence  $K$  is nilpotent. Therefore Theorem 3 follows from Theorem 1.

Suppose  $p \neq 2$ . Define an integer  $r \geq 0$  by the condition  $G(I^r) \neq \{1_n\}$ ,  $G(I^{r+1}) = \{1_n\}$ . Suppose  $r \geq 1$ ; then by Lemma 5,  $V_1(1, K/\mathbb{Q})$  acts trivially on  $G(I^r)$ . Hence  $G(I^r) \subset GL_n(O_F)$ , where  $F$  is the subfield of  $K$  corresponding to  $V_1(I, K/\mathbb{Q})$ , which is normal in  $V_0(I, K/\mathbb{Q})$ . Since  $V_0(I, K/\mathbb{Q})/V_1(I, K/\mathbb{Q})$  is cyclic,  $F$  is abelian over  $\mathbb{Q}$ . Since  $p$  is totally ramified in  $K$ ,  $p$  is also totally ramified in  $F$  and  $G(F/\mathbb{Q}) = V_0(I, K/\mathbb{Q})/V_1(I, K/\mathbb{Q})$  acts on  $G(I^r) (\subset GL_n(O_F))$ . By Theorem 1, we have  $G(I^r) \subset G(\mathbb{Z})$  and hence by Lemma 2 we have  $G(I^r) = \{1_n\}$ , which is the contradiction. Thus we have  $r = 0$  and hence  $G(I) = \{1_n\}$ . For  $g \in G$  and  $u \in G(K/\mathbb{Q}) = V_0(I, K/\mathbb{Q})$ , we have  $u(g) \equiv g \pmod{I}$  and hence  $u(g)g^{-1} \in G(I) = \{1_n\}$ . Thus  $G(K/\mathbb{Q})$  acts trivially on  $G$ . ■

## References

- [1] Bartels H -J and Kitaoka Y, Endliche arithmetische Untergruppen der  $GL_n$ , *Reine Angew. Math.* 313 (1980), 151–156
- [2] Kitaoka Y, Finite arithmetic subgroups of  $GL_n$ , II, *Nagoya Math. J.* 77 (1980) 137–143
- [3] Kitaoka Y, *Arithmetic of quadratic forms*, (Cambridge: Cambridge University Press) 1993

# Reduction theory over global fields

T A SPRINGER

Mathematisch Instituut, Rijksuniversiteit Utrecht, Budapestlaan 6, Postbus 80.010, 3508 TA Utrecht, Netherlands

Dedicated to the memory of Professor K G Ramanathan

**Abstract.** The paper contains an exposition of the basic results on reduction theory in reductive groups over global fields, in the adelic language. The treatment is uniform: number fields and function fields are on an equal footing.

**Keywords.** Reduction theory; global fields; number fields; function fields.

## Introduction

The basic results on reduction theory for a linear algebraic group  $G$  over the field of rational numbers were established by Borel and Harish-Chandra in [3]. One of these results is the construction of a fundamental set for an arithmetic subgroup  $\Gamma$  of the real Lie group  $G(\mathbf{R})$ . For another one, the criterion for compactness of the quotient  $G(\mathbf{R})/\Gamma$ , a more direct method of proof was given by Mostow and Tamagawa [8]. Godement and Weil [5] showed that this method can also be used to obtain fundamental sets. They used the language of adèles.

Reduction theory for linear algebraic groups over number fields is reduced to groups over  $\mathbf{Q}$  by restriction of the ground field. For groups over global fields of positive characteristic, i.e. function fields of dimension one over a finite field of constants, the method of Mostow and Tamagawa can also be used, but only under some restrictions on the characteristic (see [1]). Using another method, involving the study of semi-simple group schemes over complete curves, Harder [6] proved the basic results over function fields without restrictions on the characteristic.

Some 25 years ago, in unpublished seminar notes, I tried to give a uniform treatment of the reduction theory over global fields, by the method of [5], also using Harder's idea to employ Galois descent. This attempt was not successful; there was a gap in the notes. However, they contain a proof of the compactness theorem.

In the meantime, no uniform treatment of the basic results on reduction theory seems to have appeared in the literature. The present note, which is to a large extent expository, attempts to give such a treatment. The method is essentially that of the old notes. But I have abandoned the method of Mostow and Tamagawa altogether. Galois descent is used instead.

In applying the method of [8] (and its extension in [5]) one encounters a somewhat subtle question. This method seems to involve an application of the following strong version of the Hilbert-Mumford theorem. Let  $G$  be a reductive group over a field  $k$ , acting linearly in a vector space  $V$ , everything being defined over  $k$ . Let  $\xi \in V(k)$  be a

non-zero instable vector. Then there is a cocharacter of  $G$  which is defined over  $k$  such that  $\xi$  is instable for the corresponding  $G_m$ -action.

One knows that this result holds if  $k$  is perfect. But there are counter examples for non-perfect fields (see [7, 5, 6]). This might explain that application of the method of Mostow and Tamagawa in the case of arbitrary function fields has not been successful.

## 1. Preliminaries

1.1. In the sequel  $k$  denotes a global field, i.e. either a finite extension of  $\mathbb{Q}$  or a function field of dimension one with a finite field of constants. In the latter case  $k$  is a finite separable extension of a purely transcendental extension  $k_0$  of a finite field, of transcendence degree one. In the first case we put  $k_0 = \mathbb{Q}$ .

Let  $G$  be a  $k$ -group, i.e. a linear algebraic group which is defined over  $k$ . We refer to [2] for the theory of linear algebraic groups. Denote by  $G(k)$  the group of  $k$ -rational points of  $G$  and by  $G(A, k)$  or  $G(A)$  the corresponding adèle group. It is a locally compact group, containing  $G(k)$  as a discrete subgroup (see [10, Ch. 1] for the basic results on adèles).

We denote by  $G(A, k)^0$  or  $G(A)^0$  the closed subgroup of  $G(A, k)$  consisting of the adèles  $g$  such that for each rational character  $\chi$  of  $G$  which is defined over  $k$  the idele norm  $|\chi(g)|$  equals one. Then  $G(k)$  is a subgroup of  $G(A, k)^0$ . If  $G$  is the multiplicative group then  $G(A, k)$  is the group  $I(k)$  of ideles and  $G(A, k)^0$  is the group  $I(k)^0$  of ideles of norm one. We denote by  $C(G, k)$  or  $C(G)$  the quotient space  $G(A, k)^0/G(k)$ . It is well-known that  $I(k)^0/k^*$  is compact.

If  $\phi: G \rightarrow H$  is a homomorphism of  $k$ -groups we denote by  $\tilde{\phi}$  the induced homomorphism of topological spaces  $C(G, k) \rightarrow C(H, k)$ .

1.2. We now review some auxiliary results. Let  $H$  be a closed  $k$ -subgroup of  $G$  and let  $i$  be the injection  $H \rightarrow G$ . We say that a triple  $(V, \rho, \xi)$  is a  $k$ -representation of  $G$  adapted to  $H$  if  $V$  is a vector space over  $k$  (in the sense of algebraic geometry),  $\rho: G \rightarrow GL(V)$  is a  $k$ -representation of  $G$  and  $\xi \in V(k)$  is a non-zero vector such that  $H$  is the stabilizer of the line  $L$  through  $\xi$  for the  $G$ -action in  $V$  defined by  $\rho$ . It is known that such a triple exists. We may assume that  $\rho$  induces a  $k$ -isomorphism of the quotient space  $G/H$  onto the  $G$ -orbit of  $L$  in the projective space  $\mathbb{P}(V)$  (see [2, 5.1, 6.8]).

1.3. *Lemma.*  $\tilde{i}$  is a homeomorphism of  $C(H, k)$  onto a closed subspace of  $C(G, k)$ .

It is clear that  $\tilde{i}$  is an injective continuous map. Let  $(V, \rho, \xi)$  be as above. Then  $H(A)^0 \cdot G(k)$  is the inverse image of  $I(k)^0 \rho(G(k)) \cdot \xi$  under the continuous map  $g \mapsto \rho(g)^{-1}v$  of  $G(A)$  to  $V(A)$ . It suffices to prove that the induced map  $i': X/H(A)^0 \cdot (G(k)/G(k)) \rightarrow G(A)/G(k)$  is a homeomorphism onto a closed subspace. Since  $I(k)^0$  is the product of  $k^*$  and a compact set and since  $k^* \rho(G(k))v$  is discrete in  $V(A)$  we have that  $I(k)^0 \rho(G(k)) \cdot \xi$  is closed. Hence  $H(A)^0 \cdot G(k)$  is closed in  $G(A)$ , and the image of  $i'$  is closed. Since  $i'$  is open (see for example [10, p. 28–29]) it is a homeomorphism.

### 1.4. Restriction of the ground field

Let  $l$  be a finite separable extension of  $k$ . If  $G$  is an  $l$ -group denote by  $H = \Pi_{l/k} G$  the  $k$ -group obtained from  $G$  by restricting the ground field to  $k$  ([10, 1.3]). Denote by

$\phi_G: H \rightarrow G$  the canonical  $l$ -morphism. If  $G$  is a  $k$ -group we also have a  $k$ -homomorphism  $\psi_G: G \rightarrow H$  with  $\phi_G \circ \psi_G = id$ . We have an isomorphism of topological groups

$$H(A, k) \rightarrow G(A, l), \quad (1)$$

inducing an isomorphism  $H(k) \simeq G(l)$ .

If  $G$  is a  $k$ -group, the composite of  $\psi_G$  and the isomorphism (1) is the canonical injection  $G(A, k) \rightarrow G(A, l)$ .

**1.5. Lemma.** *Let  $l$  be a finite separable extension of  $k$ . The canonical injection  $C(G, k) \rightarrow C(G, l)$  is a homeomorphism onto a closed subspace.*

This follows from the preceding observations and 1.3.

Let  $H$  be a normal  $k$ -subgroup of the  $k$ -group  $G$  and let  $\pi: G \rightarrow G/H$  be the canonical homomorphism. A section over  $k$  for  $\pi$  is a  $k$ -morphism  $\sigma: G/H \rightarrow G$  such that  $\pi \circ \sigma = id$ . Then  $G(l) \rightarrow (G/H)(l)$  is surjective for any  $k$ -algebra  $l$ .

**1.6. Lemma.** *Assume that  $\sigma$  is a section over  $k$ .*

- (i) *The canonical homomorphism  $G(A) \rightarrow (G/H)(A)$  is surjective;*
- (ii) *Assume that the group of  $k$ -characters of  $H$  is trivial and that both  $C(H)$  and  $C(G/H)$  are compact. Then  $C(G)$  is compact.*

We skip the easy proof of (i). In the situation of (ii) let  $K$  and  $K'$  be compact sets in  $(G/H)(A)$  and  $H(A)$  such that  $(G/H)(A) = K \cdot (G/H)(k)$  and  $H(A) = K' \cdot H(k)$ , respectively. Then  $G(A)^0$  is a closed subset of  $\sigma(K) \cdot K' \cdot G(k)$ , and the assertion follows.

A special case where sections exist, is when  $G$  is the semi-direct product over  $k$  of  $H$  and a  $k$ -subgroup  $L$ . In that case  $G(A, k)$  is the semi-direct product of  $H(A, k)$  and  $L(A, k)$ .

## 1.7. PROPOSITION

*Let  $G$  be a connected solvable  $k$ -group which is split over  $k$ . Then  $C(G, k)$  is compact.*

For split groups see [2, § 15]. The proposition is well-known if  $G = G_a$  or  $G_m$ . In the general case  $G$  is the semi-direct product of a  $k$ -split maximal torus and its unipotent radical, which is also  $k$ -split. By the lemma the proof is reduced to the case that  $G$  is either a torus or a unipotent group. The first case reduces to the case of  $G_m$ . In the second case we have a normal sequence of connected split  $k$ -subgroups such that the successive quotients are all  $k$ -isomorphic to  $G_a$ . Since for any connected normal  $k$ -subgroup of our group  $G$  sections exist (by a result of Rosenlicht, see [9, Th. 1]), the lemma reduces this case to  $G_a$ .

## 1.8. Heights

Let  $V$  be a finite dimensional vector space over  $k$ , in the sense of algebraic geometry. Denote by  $V(A)$  the corresponding adele space and by  $GL(V, A)$  the group of invertible automorphisms of the  $A$ -module  $V(A)$ . This group is isomorphic to the adele group  $GL(V)(A)$ . We transport the structure of topological group of the latter group to  $GL(V, A)$ .

We say that  $x \in V(A)$  is *primitive* if there is  $g \in GL(V, A)$  such that  $g \cdot x$  is a non-zero

the norm  $\|x\|_v$ , equals the maximum of the absolute values of  $x$  with respect to a fixed basis of  $V(k)$  (the same then holds for any other basis, for almost all  $v$ ). For  $x \in V(A)$  primitive we put

$$\|x\| = \prod_v \|x_v\|_v. \quad (2)$$

We call such a function on the set of primitive elements of  $V(A)$  a *height*.

We list some properties of a height  $\|\cdot\|$ .

- (a) for all ideles  $t$  and all primitive  $x \in V(A)$  we have  $\|t.x\| = |t| \|x\|$ ;
- (b) if  $\|\cdot\|'$  is another height, the ratio  $\|x\|^{-1} \|x\|'$  lies in a fixed compact subset of  $\mathbf{R}_+^*$ , where  $x \in V(A)$  is primitive;
- (c) if  $(x_n)$  is a sequence of primitive vectors which converges to 0 in  $V(A)$  then  $\|x_n\|$  tends to 0 in  $\mathbf{R}$ ;
- (d) If  $(x_n)$  is a sequence of primitive vectors such that  $\|x_n\|$  tends to 0 in  $\mathbf{R}$  then there exist  $\lambda_n$  in  $k$  such that the sequence  $(\lambda_n x_n)$  converges to 0 in  $V(A)$ .

This is well-known (see [5, 1.1]). (a) and (b) are easy. To prove (c) it suffices to consider the case that in (2) we have for all places  $v$  that  $\|x_v\|$  is the maximum of the absolute values of coordinates with respect to a given basis of  $V(k)$ . Let

$$K = \{x \in V(A) \mid \|x_v\| = 1 \text{ for all } v\}.$$

This is a compact set and for any primitive  $x$  there is an idele  $t$  such that  $t.x \in K$ . Using this fact, the proof of (c) is straightforward and the proof of (d) reduces to the case that  $V$  has dimension one, which is well-known.

### 1.9. Reduction theory for $GL(2)$

We now take  $k = k_0$ . Let  $V$  be the standard 2-dimensional vector space. So  $V(k) = k^2$ ,  $V(A) = A^2$ ,  $GL(V, A) = GL(2, A)$ , where  $A = A(k_0)$ . We use a particular height. If  $k_0 = \mathbf{Q}$  we define for a finite place  $v$   $\|x_v\|$  to be the maximum of the absolute values of the coordinates with respect to the canonical basis  $(e_i)$ . For the infinite place  $v$  of  $\mathbf{Q}$  it is the Euclidean length of  $x_v$ . If  $k_0$  is a function field we take  $\|x_v\|$  to be the maximum of the absolute values of these coordinates, for all places.

For all places  $v$  the subgroup  $M_v$  of  $GL(2, k_v)$  preserving  $\|\cdot\|$  is compact. So  $M = \prod_v M_v$  is a compact subgroup of  $GL(2, A)$ . It is well-known that any  $g \in GL(2, A)$  can be written in the form  $g = m.t$ , with  $m \in M$  and  $t$  upper triangular. Denote the first and last diagonal ideles of such a  $t$  by  $t_1, t_2$ . For any  $c > 0$  let  $T(c)$  be the set of upper triangular elements  $t$  in  $GL(2, A)$  with  $|t_1/t_2| \leq c$ . The next result goes back to Gauss.

### 1.10 PROPOSITION

*There exists a constant  $c > 0$  such that  $GL(2, A) = M.T(c).GL(2, k)$ .*

The proof will show that we may take  $c = 2/\sqrt{3}$ . Let  $g \in GL(2, A)$ . We have to find  $\gamma \in GL(2, k)$  such that  $g\gamma \in M.T(c)$ . Since  $V(k)$  is discrete in  $V(A)$  it follows from property (d) of heights that the set of numbers  $\|g.\xi\|$  where  $\xi$  runs through the non-zero vectors of  $V(k)$ , is bounded away from zero. We may therefore assume that  $\|g.e_1\| \leq \|g.\xi\|$  for all such  $\xi$ . Put  $g = m.t$ , as before. Then  $g.e_1 = t_1$  and for all  $\lambda, \mu \in k$ , not both zero,



and  $u \in A$  we have

$$|t_1| \leq \|(\lambda + \mu u)t_1 e_1 + \mu t_2 e_2\|.$$

Put  $x = t_1/t_2$ . Then

$$|x| \leq \|(u + v)xe_1 + e_2\|, \quad (3)$$

for all  $v \in k$ .

If  $k = \mathbf{Q}$  we multiply  $x$  by an element of  $k^*$  (modifying  $u$  and  $v$ ) such as to obtain an idele whose components at the finite places are all 1. Then take  $v \in k$  such that  $\|u_v + v\| \leq 1$  for all finite  $v$  and  $\leq 1/2$  at the infinite place. Now (3) gives

$$|x| \leq \sqrt{(1 + |x|^2)/4},$$

whence  $|x| \leq 2/\sqrt{3}$ .

If  $k$  is a field of rational functions in one indeterminate over a finite field with  $q$  elements we take as infinite place the obvious one. Proceeding as before we see that we can now find  $v$  such that even  $\|u_v + \mu\| \leq q^{-1}$  at the infinite place. The inequality (3) gives  $|x| \leq 1$ . We have a bound as required for  $t_1 t_2^{-1}$ . The same argument works for  $SL(2)$  and gives the following.

### 1.11. COROLLARY.

*There exists a compact subgroup  $M'$  of  $SL(2, A)$  and a constant  $c > 0$  such that  $SL(2, A) = M' \cdot (T(c) \cap SL(2, A)) \cdot SL(2, k)$ .*

## 2. Reduction theory

### 2.1. Statement of the main results

We now assume that  $G$  is a connected reductive  $k$ -group. We fix a minimal parabolic  $k$ -subgroup  $P$  of  $G$ . Let  $U$  be its unipotent radical. It is a  $k$ -split unipotent group. We also fix a maximal  $k$ -split torus  $S$  of  $G$  which lies in  $P$ . The centralizer  $L$  of  $S$  is a  $k$ -Levi group of  $P$  and  $P$  is the semi-direct product over  $k$  of  $L$  and  $U$ . We denote by  $R$  the root system of  $(G, S)$ , by  $R^+$  the set of positive roots defined by  $P$  and by  $\Delta$  the basis of  $R$  defined by  $R^+$ . For the basic facts on reductive  $k$ -groups we refer to [4].

The homogeneous space  $G/P$  is a projective  $k$ -variety. It follows from [loc.cit., 4.13] that the canonical map  $G(A)/P(A) \rightarrow (G/P)(A)$  is a homeomorphism. It follows that  $G(A)/P(A)$  is compact.

Let  $X(P)$  be the group of  $k$ -characters of  $P$ . We have a homomorphism  $P(A) \rightarrow \text{Hom}(X(P), \mathbf{R}_+^*)$  whose kernel is  $P(A)^0$ , and there is a similar homomorphism for  $S$ . One knows that restriction of characters identifies  $X(P)$  with a subgroup of finite index of  $X(S)$ . It then readily follows that there is a finite subset  $F$  of  $P(A)$  such that  $P(A) = F \cdot S(A) \cdot P(A)^0$ .

We conclude that there is a compact subset  $K$  of  $G(A)$  such that

$$G(A)^0 = K \cdot (S(A) \cap G(A)^0) \cdot P(A)^0.$$

If  $c$  is a strictly positive constant we define  $S(c)$  to be the set of  $s \in (S(A) \cap G(A)^0)$  such

that  $|\alpha(s)| \leq c$  all  $\alpha \in \Delta$  and we put  $\mathcal{S}(c) = K.S(c).P(A)^0$ . The following statements are the main results of reduction theory.

(C) (Compactness theorem)  $C(G)$  is compact if and only if  $G$  is anisotropic (i.e.  $P = G$ ).

(F) (Fundamental set theorem) If  $G$  is isotropic there is  $c$  with  $G(A)^0 = \mathcal{S}(c).G(k)$ .

2.2. Remarks. (a) Since the Levi group  $L$  of  $P$  is anisotropic and the unipotent radical  $U$  is  $k$ -split, it follows from the compactness theorem that there is a compact set  $K_1$  in  $L(A)^0$  such that  $L(A)^0 = K_1.L(k)$ . The fundamental set theorem then implies that  $G(A)^0 = K.K_1.S.U(A)G(k)$ . Using 1.7 we conclude that there are compact sets  $K' \subset G(A)^0$  and  $K'' \subset U(A)$  with  $G(A)^0 = K'.T(c).K''.G(k)$ .

(b) Let  $V$  be a vector space over  $k$  such that  $G$  is a closed  $k$ -subgroup of  $GL(V)$ . Assume that  $C(G)$  is compact and let  $K_0$  be a compact subset of  $G(A)^0$  such that  $G(A)^0 = K_0.G(k)$ . Choose an open neighbourhood  $U$  of 0 in  $V(A)$  such that  $U \cap V(k) = \{0\}$  and that  $K_0^{-1}.U \subset U$ . Then  $G(A)^0.V(k) \cap U = \{0\}$ . It follows that 0 is an isolated point of  $G(A)^0.V(k)$ . On the other hand, if  $G$  is isotropic over  $k$  there exists a non-trivial  $k$ -split subtorus  $S$  of the commutator subgroup of  $G$ . Then  $S(A)$  is a subgroup of  $G(A)^0$ . Let  $\xi \in V(k)$  of  $S$  be a weight vector of  $S$  whose weight  $\chi$  cannot be extended to a character of  $G$  and choose a sequence  $(s_n)$  in  $S(A)$  such that  $\chi(s_n)$  converges to zero in  $A$ . Then  $(s_n.\xi)$  converges to 0 in  $V(A)$ . We conclude that  $G$  is anisotropic if  $C(G)$  is compact. Then proof of the converse statement is the crucial part of the proof of the compactness theorem.

### 2.3. Auxiliary results

The parabolic  $k$ -subgroups of  $G$  containing  $P$  are parametrized by the subsets of  $\Delta$ . If  $\Pi \subset \Delta$  we denote by  $P_\Pi$  the parabolic  $k$ -subgroup containing  $P$  such that the root system of the Levi group of  $P_\Pi$  which contains  $S$  has basis  $\Pi$  (so  $P_\emptyset = P$ ).

We number the elements of  $\Delta$ , say  $\Delta = \{\alpha_1, \dots, \alpha_r\}$ . For  $i \in [0, r]$  put

$$P_i = P_{\{\alpha_{i+1}, \dots, \alpha_r\}}.$$

So  $P_0 = G$ ,  $P_r = P$ . For  $i \in [1, r]$  let  $(\rho_i, V_i, \xi_i)$  be a representation adapted to  $P_i$ . Notice that for  $j \geq i$  we have that the restriction of  $\rho_j$  to  $L_i$  is adapted to the parabolic  $k$ -subgroup  $L_i \cap P_j$  of  $L_i$ .

We fix a height  $\|\cdot\|_i$  on  $V_i(A)$ . For  $c > 0$  we put

$$\mathcal{R}(c) = \{g \in G(A)^0 \mid \|\rho_i(g)\xi_i\|_i \leq c \|\rho_i(g\gamma)\xi_i\|_i \text{ for } i \in [1, r], \gamma \in P_i(k)\}.$$

It follows from property (d) of heights that  $G(A)^0 = \mathcal{R}(1).G(k)$ .

The next lemma gives a reduction to rank one, following [5, 9.3].

2.4. Lemma. Assume that (F) holds for the  $k$ -groups  $L_i$  ( $1 \leq i \leq r-1$ ).

(i) There is  $c'$  such that  $\mathcal{R}(c) \subset \mathcal{S}(c')$ ;

(ii) (F) holds for  $G$ .

(ii) follows from (i), by the remark we just made. We prove (i) by induction on the rank  $r$ . If  $r = 1$  then  $\xi_1$  is a weight vector for  $S$  whose weight is a rational multiple of the only simple root  $\alpha$ . It follows from (F) that there is a constant  $c_0$  such that for  $g \in G(A)^0$  there is  $\gamma \in G(k)$  with  $g\gamma = xsy$ , where  $x \in K$ ,  $s \in S(A) \cap G(A)^0$ ,  $y \in P(A)^0$  and

$|\alpha(s)| \leq c_0$ . Hence  $\inf_{g \in G(k)} \|\rho_1(g) \cdot \xi_1\|_1 \leq c_0$ . It follows that if  $g = xsy \in \mathcal{R}(c)$  we have  $\|g \cdot \xi_1\|_1 \leq c_0$ , from which one concludes that  $|\alpha(s)|$  must be bounded by a constant.

Now let  $r > 1$  and take  $g = xsy \in \mathcal{R}(c)$ . We can write  $y = y_1 u_1$ , where  $y_1 \in L_1(A)^0$ ,  $u_1 \in U_1(A)$  ( $U_1$  denoting the unipotent radical of  $P_1$ ). It is immediate that  $sy_1$  lies in a set like  $\mathcal{R}(c)$  for the group  $L_1$ . By induction it follows that  $|\alpha_i(s)|$  is bounded by a constant for  $i > 1$ . A similar argument, using the rank one group  $L_{r-1}$  gives a bound for  $|\alpha_1(s)|$  and (i) follows.

2.5. For  $\alpha \in \Delta$  we denote by  $(\rho_\alpha, V_\alpha, \xi_\alpha)$  a representation adapted to the maximal parabolic subgroup  $P(\alpha) = P_{\Delta - \{\alpha\}}$ . Then  $\xi_\alpha$  is a weight vector for  $S$  whose weight  $\chi_\alpha$  is a strictly positive multiple of  $\alpha$ .

We denote by  $W$  the Weyl group  $N_G(S)/Z_G(S)$ . It is the Weyl group of the root system  $R$ . If  $\Pi \subset \Delta$  denote by  $W_\Pi$  the Weyl group of the Levi group of  $P_\Pi$  containing  $S$ , relative to  $S$ . Then  $W_\Pi$  is a parabolic subgroup of  $W$ . For  $\alpha \in \Pi$  we put  $W(\alpha) = W_{\Delta - \{\alpha\}}$ .

For  $w \in W$  denote by  $\dot{w}$  a representative lying in the group of rational points  $N_G(S)(k)$ . By Bruhat's lemma we have  $G(k) = \bigcup_{w \in W} U(k) \dot{w} P(k)$ . Fix  $\gamma \in G(k)$  and write  $\gamma = \mu \dot{w} v$  where  $\mu \in U(k)$ ,  $w \in W$ ,  $v \in P(k)$ . Put  $\mathcal{S}'(c') = K' \cdot S(c') \cdot P(A)^0$ , where  $K'$  is another compact set and  $c' > 0$ . Let  $g = xsy = x' s' y' \gamma$  with  $x \in K$ ,  $x' \in K'$ ,  $s, s' \in S(A) \cap G(A)^0$ ,  $y, y' \in P(A)^0$ , be an element of  $\mathcal{S}(c) \cap \mathcal{S}'(c') \gamma$ .

## 2.6. PROPOSITION

Assume that  $C(L)$  is compact. Let  $\alpha \in \Delta$ .

If the set of positive numbers  $|\alpha(s)|$ , where  $g = xsy = x' s' y' \gamma$  runs through  $\mathcal{S}(c) \cap \mathcal{S}'(c') \gamma$ , is not bounded away from zero then  $\gamma \in P(\alpha)$ .

This is a variant of [5, lemme 3]. Since  $C(L)$  is compact we may assume (changing  $K$  and  $K'$ ) that  $y$  and  $y'$  lie in  $U(A)$ . Then  $g' = sy(\gamma)^{-1}(s' y')^{-1}$  lies in the compact set  $K^{-1} \cdot K'$ . Let  $\beta \in \Delta$  and fix a height  $\|\cdot\|$  on  $V_\beta(A)$ . We have

$$\|\rho_\beta(g') \cdot \xi_\beta\| = |\chi_\beta(s')^{-1}| \cdot |\chi_\beta(w.s)| \|syv^{-1}s^{-1}\dot{w}^{-1} \cdot \xi_\beta\|.$$

Since  $C(U, k)$  is compact it follows from properties (c) and (d) of heights, using that  $\xi_\beta$  is a positive multiple of  $\beta$ , that  $|\beta((s')^{-1}(w.s))|$  lies in a compact set of  $\mathbf{R}_+^*$ . We conclude that there is a constant  $d$  such that for all  $\beta \in \Delta$  we have

$$|(w^{-1} \cdot \beta)(s)| \leq d |\beta(s')| \leq c' d. \quad (4)$$

Now assume that  $w \notin W(\alpha)$ . Then there is a positive root  $\delta$  such that  $w^{-1} \cdot \delta = -\sum_{\beta \in \Delta} n_\beta \beta$ , with  $n_\alpha > 0$ . It follows from (4) that there is a constant  $e$  such that

$$|(w^{-1} \cdot \delta)(s)| \leq e.$$

On the other hand we have

$$|w^{-1} \cdot \delta(s)| = \prod_{\beta \in \Delta} |\beta(s)^{-n_\beta}| \geq |\alpha(s)|^{-n_\alpha} \prod_{\beta \neq \alpha} c^{-n_\beta}.$$

Since the numbers  $n_\beta$  are bounded the last two inequalities imply that  $|\alpha(s)|$  is bounded below by a strictly positive constant. So if  $|\alpha(s)|$  is not bounded away from zero, we must have  $w \in W(\alpha)$  and  $\gamma \in P(\alpha)$ , proving the proposition.

### 3. Proof of the main results

#### 3.1. A reduction

The field  $k$  is a finite separable extension of a subfield  $k_0$  which is either  $\mathbf{Q}$  or a field of rational functions in one variable over a finite field. Let  $H = \Pi_{k/k_0} G$  be the  $k_0$ -group obtained by restriction of the ground field. We use the notations of 1.4. We have the following facts.

(a)  $\phi_G$  induces isomorphisms  $H(k_0) \simeq G(k)$ ,  $H(A, k_0) \simeq G(A, k)$ .

See 1.4. We denote these isomorphisms also by  $\phi_G$ .

(b)  $\phi_G$  induces an isomorphism  $H(A, k_0)^0 \simeq G(A, k)^0$ .

For the (easy) proof see [1, p. 14].

The maximal  $k$ -split torus  $S$  is a  $k_0$ -group. The composite of  $\psi_S$  and the canonical morphism  $\Pi_{k/k_0} S \rightarrow H$  is a  $k_0$ -homomorphism  $\mu: S \rightarrow H$ .

(c)  $S' = \mu S$  is a maximal  $k_0$ -split torus of  $H$  and there is a bijection of  $R$  onto the root system of  $(H, S')$ .

This is straightforward.

(d)  $P' = \Pi_{k/k_0} P$  is a minimal parabolic subgroup over  $k_0$  in  $H$  and

$$\phi_G(P'(A, k_0)^0) = P(A, k)^0.$$

(e) For  $c > 0$  there exists  $c'$  such that  $\phi_G(S'(c)) \subset S(c')$ .

See [4, no. 6] and [1, p. 14] for (d) and (e). The facts just stated imply that it suffices to prove the statements (C) and (F) in the case that  $k = k_0$ , which we assume from now on.

#### 3.2. Split groups

We first prove (F) in the case that  $G$  is split over  $k (= k_0)$ , i.e. that  $S$  is a maximal torus of  $G$ . In that case  $P$  is a Borel group, which is a  $k$ -split connected solvable group. By 1.7 we know that  $C(L)$  is compact. Also, the groups  $L_i$  of 2.3 are split over  $k$ . It then follows from 2.3 by an easy induction that property (F) for  $G$  is a consequence of the following lemma.

**3.3. Lemma.** *If  $G$  has semi-simple rank one and is split over  $k$  then (F) holds.*

A  $k$ -group  $G$  with these properties is  $k$ -isomorphic to a product  $H \times T$ , where  $T$  is a  $k$ -split torus and  $H$  is one of the groups  $GL(2)$ ,  $SL(2)$ ,  $PGL(2)$ . (This must be well-known, but as I do not know a reference a proof will be sketched below.) It suffices to prove the lemma for  $H$ . If  $H = PGL(2)$  there is the obvious map  $C(GL(2), k) \rightarrow C(H, k)$ , which is surjective. (Notice that for any field  $l$  the canonical map  $GL(2, l) \rightarrow PGL(2, l)$  is surjective.) A set  $\mathcal{S}(c)$  for  $GL(2)$  is mapped onto a similar set for  $H$ . So we may assume that  $H = GL(2)$  or  $SL(2)$  and then 1.10 and 1.11 establish the lemma.

The proof of the result on the structure of  $G$  uses the root datum of  $(G, S)$ , say  $(X, X^\vee, R, R^\vee)$ . Here  $X$  is the character group of  $S$ ,  $R = \{\pm \alpha\}$  the root system,  $X^\vee$  the dual of  $X$  and  $R^\vee = \{\pm \alpha^\vee\}$  the dual root system. If  $\alpha = 2\chi \in 2X$  then  $X$  is the direct sum of  $Z\chi$  and  $(\alpha^\vee)^\perp$  and  $G$  is  $k$ -isomorphic to the direct product of  $SL_2$  and a torus. Similarly, if  $\alpha^\vee \in 2X^\vee$  then  $G$  is isomorphic to  $PGL_2$  times a torus. If  $\alpha \notin 2X$ ,  $\alpha^\vee \notin 2X^\vee$  choose  $\lambda \in X^\vee$  with  $\langle \alpha, \lambda \rangle = 1$ . Put  $X_0^\vee = Z\alpha^\vee + Z\lambda$ . Then  $X$  is the direct

3.4. Now assume that  $G$  is arbitrary. We fix a maximal torus  $T$  of  $G$  which is defined over  $k$  and contains  $S$ . We also fix a finite separable Galois extension  $l$  of  $k$  which splits  $T$ . Denote by  $\Gamma$  the Galois group of  $l$  over  $k$ . We denote by  $\tilde{R}$  the root system of  $(G, T)$  and by  $\tilde{W} = N_G(T)/T$  its Weyl group. If  $w \in \tilde{W}$  we denote by  $\dot{w} \in \tilde{W}$  a representative in  $G(l)$ . The roots of the relative root system  $R$  are the non-trivial restrictions to  $S$  of the roots of  $\tilde{R}$ . We fix a system of positive roots  $\tilde{R}^+$  such that the roots in  $R^+$  are restrictions of roots in  $\tilde{R}^+$ . Let  $\tilde{\Delta}$  be the basis defined by  $\tilde{R}^+$ . If  $\Pi \subset \tilde{\Delta}$  we denote by  $\tilde{P}_\Pi$  the corresponding parabolic  $l$ -subgroup.

Since  $T$  is split over  $l$ , all characters of  $T$  are defined over  $l$ . Hence the Galois group  $\Gamma$  acts on  $\tilde{R}$ . Let  $B \supset P$  be the Borel subgroup defined by  $\tilde{R}^+$  and let  $\tilde{U}$  be its unipotent radical. For  $s \in \Gamma$  there is  $w_s \in \tilde{W}$  such that  $s.B = \text{Int}(\dot{w}_s)B$ . There is an action  $\iota$  of  $\Gamma$  on  $\tilde{\Delta}$  such that for  $s \in \Gamma$ ,  $\alpha \in \tilde{\Delta}$  we have  $s.\alpha = w_s(L(s).\alpha)$ . We then have  $s.P_\Pi = \text{Int}(\dot{w}_s)P_{\iota(s).\Pi}$  if  $\Pi \subset \tilde{\Delta}$ .

3.5. *Proof of (C).* As we remarked in 2.2 the burden of the proof of (C) is to show that  $C(G, k)$  is compact if  $G$  is anisotropic. This we now assume. We identify  $G(A, k)^0$  with a closed subgroup of  $G(A, l)^0$ . The Galois group  $\Gamma$  acts on the latter group, and  $G(A, k)^0$  is the set of elements fixed by all of  $\Gamma$ . Since  $G$  is  $l$ -split, we know that (C) and (F) hold over  $l$ . Put  $\mathcal{S}'(c) = K.T(c).\tilde{U}(A, l)$ , where  $K$  is a compact set. We assume that  $G(A, l)^0 = \mathcal{S}'(c).G(l)$ . Take  $g \in G(A, k)^0$ . There are  $x \in K$ ,  $t \in T(c)$ ,  $u \in U(A, l)$ ,  $\gamma \in G(l)$  with  $g = xtu\gamma$ . For all  $s \in \Gamma$  we have  $s(xtu\gamma) = xtu\gamma$ . This can be rewritten as

$$((s.x)\dot{w}_s)(w_s^{-1}(s.t)).u' = xtu(\gamma(s.\gamma)^{-1}\dot{w}_s),$$

where  $u' = \text{Int}(\dot{w}_s^{-1})(s.u) \in U$ . Let  $\alpha \in \tilde{\Delta}$ . It follows from 2.6 that there is a constant  $d$  such that if  $|\alpha(t)| \leq d$  we have  $\gamma(s.\gamma)^{-1}\dot{w}_s \in \tilde{P}(\alpha)$ , for all  $s \in \Gamma$ . Then  $s.\text{Int}(\gamma^{-1})\tilde{P}(\iota(s)^{-1}.\alpha) = \text{Int}((s.\gamma)^{-1}\dot{w}_s)\tilde{P}(\alpha) = \text{Int}(\gamma^{-1})\tilde{P}(\alpha)$  and the proper parabolic subgroup

$$Q = \bigcap_{s \in \Gamma} \text{Int}(\gamma^{-1})\tilde{P}(\iota(s).\alpha)$$

is  $\Gamma$ -stable, hence is defined over  $k$ . But this is impossible if  $G$  is anisotropic. It follows that for all  $\alpha \in \tilde{\Delta}$  we have  $|\alpha(t)| \geq d$ . Since  $U(A, l)/U(l)$  is compact we conclude that the image of  $G(A, k)^0$  in  $C(G, l)$  is relatively compact. By 1.2 we have that  $C(G, k)$  is compact and (C) follows.

3.6. *Proof of (F).* Let  $G$  be arbitrary. By (C) we know now that  $C(G, l)$  is compact. Application of 2.4 shows that it suffices to establish (F) in the case the  $G$  has semi-simple rank one. This we now assume. We proceed as in 3.5 and use the notations introduced there. In the present case, the parabolic  $k$ -subgroup  $Q$  must be  $k$ -conjugate to  $P$ . This means that, after multiplying  $g$  on the right by an element of  $G(k)$ , we may assume that  $Q = P$ . But it is known (see [4, no. 6]) that there is a unique  $\Gamma$ -orbit  $\mathcal{O}$  in  $\tilde{\Delta}$  such that  $P$  is the intersection of the  $\tilde{P}(\alpha)$  with  $\alpha \in \mathcal{O}$ . The roots in  $\mathcal{O}$  restrict to the simple

root of  $\Delta$ . We conclude that the  $\alpha \in \tilde{\Delta}$  such that  $|\alpha(t)|$  is not bounded away from zero lie in  $\mathcal{O}$  and also that if  $Q = P$  the element  $\gamma$  must lie in  $P(l)$ .

Let  $(\rho, V, \xi)$  be a  $k$ -representation of  $G$  adapted to  $P$ . Then  $\xi$  is a weight vector for the maximal  $k$ -split torus  $S$ , whose weight  $\chi$  is a positive multiple of the only simple root  $\alpha$  of the root system  $R$ . Fix a height  $\|\cdot\|$  on  $V(A, k)$ . Let  $g \in G(A, k)^0$  and write  $g = ysz$ , where  $s \in S(A, k) \cap G(A, k)^0$ ,  $z \in P(A, k)^0$  and  $y$  lies in a fixed compact set (see 2.1). Then

$$\|\rho(g) \cdot \xi\| = |\chi(s)| \|\rho(y) \cdot \xi\|.$$

By property (b) of heights we conclude that  $|\chi(s)|^{-1} \|\rho(g) \cdot \xi\|$  lies in a compact subset of  $\mathbf{R}_+^*$ .

On the other hand we can view  $g$  as an element of  $G(A, l)$ . As in 3.5 we write  $g = xtw\gamma$ . We may assume that  $\gamma \in P(l)$ . Now  $(\rho, V, \xi)$  obviously is an  $l$ -representation of  $G$  adapted to  $P$ , and  $\xi$  is a highest weight vector for that representation (relative to the Borel group  $B$ ). Its weight  $\psi$  is a linear combination of the roots in  $\tilde{\Delta}$ , with non-negative rational coefficients. It follows that for  $t \in T(c)$  the numbers  $|\psi(t)|$  lie in a compact set of  $\mathbf{R}$ . Take a height on  $V(A, l)$ . We denote it by  $\|\cdot\|$  and we assume (as we may) that its restriction to  $V(A, k)$  is the previous height. Then

$$\|\rho(g) \cdot \xi\| = |\psi(t)| \|\rho(x) \cdot \xi\|,$$

and we see that  $\|\rho(g) \cdot \xi\|$  lies in a bounded set. Then the same holds for  $|\chi(s)|$  and  $|\alpha(s)|$ , which means that  $g$  lies in a set  $\mathcal{S}(c')$ . It follows that  $G(A, k) = \mathcal{S}(c') \cdot G(k)$ , which we had to prove.

## References

- [1] Behr H, Endliche Erzeugbarkeit arithmetischer Gruppen über Funktionenkörpern, *Invent. Math.* **7** (1969) 1–32
- [2] Borel A, *Linear algebraic groups*, 2nd ed. (Springer), (1991)
- [3] Borel A and Harish-Chandra, Arithmetic subgroups of algebraic groups, *Ann. of Math.* **75** (1962) 485–535
- [4] Borel A and Tits J, Groupes réductifs, *Publ. Math. IHES* **27** (1965) 55–150
- [5] Godement R, Domaines fondamentaux des groupes arithmétiques, *Sém. Bourbaki* no. **237** (1962/63)
- [6] Harder G, Minkowskische Reduktionstheorie über Funktionenkörpern, *Invent Math.* **7** (1969) 33–54
- [7] Hesselink WH, Uniform instability in reductive groups, *J. Reine Angew. Math.* **303/304** (1978) 74–96
- [8] Mostow G D and Tamagawa T, On the compactness of arithmetically defined homogeneous spaces, *Ann. of Math.* **76** (1962) 446–463
- [9] Rosenlicht M, Questions of rationality for solvable algebraic groups over nonperfect fields, *Ann. Mat. Pura Appl.* **61** (1963) 97–120
- [10] Weil A, *Adèles and algebraic groups*, (Birkhäuser), (1982)

# Symplectic structures on locally compact abelian groups and polarizations

R RANGA RAO

Department of Mathematics, University of Illinois, 1409 West Green Street,  
Urbana, Illinois 61801, USA

Dedicated to the memory of Professor K G Ramanathan

**Abstract.** Let  $X$  be a locally compact abelian group and  $\omega(\cdot, \cdot)$  a symplectic structure on it. A polarization for  $(X, \omega)$  is a pair of totally isotropic closed subgroups  $G, G^*$  of  $X$  such that  $X = G \cdot G^*$  and  $\omega(\cdot, \cdot)$  defines a dual pairing of  $G$  and  $G^*$ . In this paper we describe a class of such groups which always admit a polarization and also discuss their structure.

**Keywords.** Symplectic structures.

## 1. Introduction

Let  $\mathcal{L}$  denote the class of locally compact Hausdorff, abelian and second countable groups. For  $X \in \mathcal{L}$ , consider an alternating bicharacter  $\omega$  on  $X$ , i.e. (i) For each  $x, y \rightarrow \omega(x, y)$  is character in  $y \in X$ , and for each  $y, x \rightarrow \omega(x, y)$  is character in  $x$  and (ii)  $\omega$  is alternating i.e.,  $\omega(x, x) = 1$  for each  $x \in X$ . Such  $\omega$  is known to provide a classification of central extensions of  $X$  by the circle group  $T$ . For the central extension  $1 \rightarrow T \rightarrow E_\omega \rightarrow X \rightarrow 1$  corresponding to the classifying invariant  $\omega$ , the analogue of the Stone-von Neumann theorem has been proved under two sets of assumptions. The first one is

$$(X, \omega) \text{ is nondegenerate} \tag{1}$$

i.e.,  $\omega(x_0, y) = 1$  for all  $y$  implies  $x_0 = 0$  (the identity element of  $X$ ) and if  $\chi$  is any continuous character of  $X$ , then there exists an  $x_0 \in X$  such that  $\chi(y) = \omega(x_0, y)$  for all  $y$ . This assumption makes it possible to apply Mackey's theory of systems of imprimitivity to  $E_\omega$  to get Stone-von Neumann theorem (see for example [M]). On the other hand, Weil [W1] used a different kind of assumption on  $(X, \omega)$  to obtain the same result. This assumption may be succinctly described by saying that  $(X, \omega)$  admits a polarization i.e., there exist closed subgroups  $G, G^*$  of  $X$  with the following properties:

- (i)  $G \cap G^* = \langle 1 \rangle$ ,  $X = G \cdot G^*$
- (ii)  $\omega(G, G) = 1$ ,  $\omega(G^*, G^*) = 1$  i.e.,  $\omega(x, y) = 1$  whenever  $x, y \in G$ , and also whenever  $x, y \in G^*$
- (iii) the mapping  $x, y \rightarrow \omega(x, y)$  of  $G \times G^*$  into  $T$  is a dual pairing of  $G$  and  $G^*$ .

Note the assumption (i) means that  $X$  is a direct sum (or direct product) of the groups  $G$  and  $G^*$  and the assumption (iii) identifies  $G^*$  as the character group of  $G$ .

Conversely if  $G \in \mathcal{L}$  and  $G^*$  is its dual group, then  $(G \times G^*, \omega)$  is nondegenerate, where  $\omega(x, x^*) = x^*(x)$  and  $\omega(\cdot, \cdot)$  in general is defined by (2). The problem we are concerned about in this paper is to find the class of groups  $X$  such that any nondegenerate symplectic structure  $\omega$  on  $X$  will admit a polarization. The main result is.

**Theorem 1.** *Let  $\mathcal{L}_0$  denote the class of groups  $X \in \mathcal{L}$  possessing the following properties:*

- (i) *The maximal compact connected subgroups of both  $X$  and its dual  $X^*$  are tori*
- (ii)  *$\dim_{\mathbb{Q}_p} \text{hom}(X, \mathbb{Q}_p) < \infty$ , for all primes  $p$*
- (iii) *The subgroup  $\{x \in X : p \cdot x = 0\}$  is finite for each prime  $p$ .*

*If  $X \in \mathcal{L}_0$  and  $\omega$  a nondegenerate symplectic structure on  $X$ , then  $(X, \omega)$  admits a polarization.*

A proof of this theorem is given in §3. In §2, we recall certain basic results on structure of groups in  $\mathcal{L}$  and also discuss the structure of groups in class  $\mathcal{L}_0$  in some detail. In §3 we consider symplectic structures  $(X, \omega)$  and prove various decomposition theorems, leading to a proof of Theorem 1. In a sequel we will discuss applications of these to the metaplectic representation.

## 2. The structure of the class of groups in $\mathcal{L}_0$

Let  $\mathcal{L}$  denote the class of second countable, Hausdorff, locally compact abelian groups. For any  $G \in \mathcal{L}$ , let  $G^* = \text{hom}(G, T)$  be the character group of  $G$ . Let  $G^0$  be the connected component of  $G$ , and let  ${}^0G = \{x \in G : x \text{ is a compact element of } G, \text{ i.e., the subgroup generated by } x \text{ has compact closure}\}$ . Then we have.

**Lemma 2.** *For any  $G \in \mathcal{L}$ ,  ${}^0G$  is a closed subgroup and it is the annihilator in  $G$  of the connected component of  $G^*$ . Moreover*

$${}^0G = \cap \{\ker \xi : \xi \in \text{hom}(G, R)\}.$$

(For proof see [HR] p. 382 and p. 390).

**Lemma 3.** *For any  $G$  in  $\mathcal{L}$ , let  $K = {}^0G \cap {}^0G$ —the subgroup of compact elements in  $G^0$ . Then  $K$  is compact, connected and is the maximal compact connected subgroup of  $G$ . Moreover  $G^0 \simeq R^n$ .  $K$  for some  $n$ . (See [HR] page 95).*

**Lemma 4.** *The subgroup  ${}^0G \cdot G^0$  is open in  $G$  and its annihilator in  $G^*$  is the maximal compact connected subgroup of  $G^*$ . In particular  $G/({}^0G \cdot G^0)$  is discrete and torsion free. ([HR], §9.26(a), p. 103).*

Using these we can next prove.

## PROPOSITION 5

*For a group  $G$  in  $\mathcal{L}$ , the following statements are equivalent:*

- (i) *the maximal compact connected subgroup of  $G^*$  is a torus, i.e., is isomorphic to  $T^m$  for some  $m$*



- (ii)  $\dim_R \text{hom}(G, R) < \infty$ , and if  $\xi \in \text{hom}(G, R)$  is such that  $\xi(G) \subseteq \mathbb{Q}$ , then either  $\xi(G) = (0)$  or  $\xi(G)$  is a lattice
- (iii) For some integers  $n$  and  $k$ ,  $G \simeq {}^0G \times R^n \cdot \mathbb{Z}^k$ . (direct sum)

*Proof.* (i)  $\rightarrow$  (iii). From Lemma 4, it follows that  $G/({}^0G \cdot G^0)$  is the character group of the maximal compact connected component of  $G^*$ . Thus  $G/({}^0G \cdot G^0) \simeq \mathbb{Z}^k$  for some  $k$ . This implies that  $G \simeq ({}^0G \cdot G^0) \cdot \mathbb{Z}^k$ . Since  $({}^0G \cdot G^0) \simeq {}^0G \cdot R^n$ , the conclusion (iii) follows.

(iii)  $\rightarrow$  (i). For (iii) implies that  $G/({}^0G \cdot G^0) \simeq \mathbb{Z}^k$  so that the maximal compact connected subgroup of  $G^*$  is  $\simeq T^k$ .

(iii)  $\rightarrow$  (ii). This is clear.

(ii)  $\rightarrow$  (i). Let  $D = G/({}^0G \cdot G^0)$ . Note the property (ii) for  $G$  implies that the same is valid for  $D$ . Let  $\dim \text{hom}(D, R) = d$  and let  $\xi_1, \dots, \xi_d$  be a basis for  $\text{hom}(D, R)$ . Consider the map  $\varphi = (\xi_1, \dots, \xi_d)$  of  $D \rightarrow R^d$ . It is then clear that the subspace spanned by  $\xi(D)$  is  $R^d$ . Let  $v_j, j = 1, 2, \dots, d$  be a basis of  $R^d$ , with  $v_j \in \xi(D)$  for all  $j$ . Let  $A$  be a  $d \times d$  matrix, nonsingular such that  $Av_j = e_j$ , where  $e_j$  is the standard basis of  $R^d$ . Let  $\eta = A\xi$ , and  $\psi = (\eta_1, \dots, \eta_d)$ . Then  $e_j \in \psi(D)$  and so  $\psi(D) \supseteq \mathbb{Z}^d$ . Also if  $\xi_j(x) = 0$  for all  $j$ , then  $x \in \ker \xi \cap \{\ker \xi : \xi \in \text{hom}(D, R)\} = {}^0D = \langle 0 \rangle$ . Thus  $\varphi$  and  $\psi$  are both injective. Let  $x_j \in D$ , be such that  $\psi(x_j) = e_j$  and let  $D_0 = \sum \mathbb{Z}x_j$ , be the subgroup generated by  $x_j, j = 1, 2, \dots, d$ . From the construction it is clear that  $D/D_0$  is torsion. For if  $x \in D$ , such that  $\pi(x) \in D/D_0$  is an element of infinite order,  $\pi$  being the canonical map  $D \rightarrow D/D_0$ , then there exists a  $\lambda \in \text{hom}(D/D_0, R)$ , such that  $\lambda(\pi(x)) \neq 0$ . Consider  $\mu = \lambda \circ \pi$ . Then  $\mu \in \text{hom}(D, R)$  and so  $\mu = \sum c_j \eta_j$ . But then  $\mu(x_j) = 0$  for all  $j$  implies that  $c_j = 0$  for all  $j$ , since  $\eta_j(x_i) = \delta_{ij}$ . Thus  $\mu = 0$ , contradicting  $\lambda(\pi(x)) \neq 0$ . Thus  $D/D_0$  is torsion. This implies that  $\mathbb{Z}^d \subseteq \psi(D) \subseteq \mathbb{Q}^d$ . Thus each  $\eta_j$  is such that  $\eta_j(D) \subseteq \mathbb{Q}$ . Since  $D$  also satisfies (ii), it follows that  $\eta_j(D)$  is a lattice. Thus there exists an integer  $N$  such that  $\mathbb{Z}^d \subseteq \psi(D) \subseteq (1/N)\mathbb{Z}^d$ . So  $\psi(D) \simeq \mathbb{Z}^d$ , or  $\psi$  being injective, it follows that  $D \simeq \mathbb{Z}^d$ . This proves that  $G$  has property (i)  $\square$

## COROLLARY 6

Let  $G \in \mathcal{L}$  be such that the maximal compact connected components of both  $G$  and  $G^*$  are tori. Then there exist closed subgroups  $G_\infty$  and  $G_f$  of  $G$  such that

- (i)  $G = G_\infty \cdot G_f$  (direct sum)
- (ii)  $G_\infty \simeq R^m \times T^n \times \mathbb{Z}^k$
- (iii)  $G_f$  is totally disconnected and every element of  $G_f$  is compact.

*Proof.* Let  $K$  be the maximal compact connected subgroup of  $G$ . Then  $K \simeq T^n$ . Since subgroups which are isomorphic to tori are direct summands, it follows that there exists a closed subgroup  $G_f$  of  ${}^0G$ , such that  ${}^0G \simeq K \cdot G_f$  direct sum. Since  $K$  is the connected component of  ${}^0G$ , it follows that  $G_f$  is totally disconnected. The rest follows from Proposition 5.

*Remark-Definition 7.* From Lemma 1, it follows that  $G$  is totally disconnected if and only if every element of  $G^*$  is compact. Thus the following statements are equivalent and the class of groups satisfying them is denoted by  $\mathcal{C}$ .

- (i)  $G$  is totally disconnected and every element of  $G$  is compact. (ii) Both  $G$  and  $G^*$  are totally disconnected. (iii) Every element of both  $G$  and  $G^*$  is compact.

To analyze the structure of this class further we recall the notion of topological  $p$ -group (see [A] or [B] for details). For  $G \in \mathcal{L}$ ,  $p$  a prime, define  $G_p = \{x \in G :$

some groups such as  $T$  we use the multiplicative notation.) Clearly  $G_p$  is a subgroup (not in general closed) of  $G$  and is called the topological  $p$ -primary component of  $G$ . The group is said to be  $p$ -primary if  $G = G_p$ . For example it is known that  $T_p =$  the  $p$ -primary component of  $T$  is the subgroup  $= \{\exp 2\pi i(m/p^n): m, n \in \mathbb{Z}\}$  or  $T_p \simeq \mathbb{Z}[1/p]/\mathbb{Z}$  as a subgroup. Let  $\mathbb{Q}_p$  be the field of  $p$ -adic numbers and  $\mathbb{Z}_p$ —the ring of  $p$ -adic integers. Then  $\mathbb{Z}_p$  is compact and its character group is isomorphic to  $T_p$  (with discrete topology).

**Lemma 8** (i) An element  $x \in G_p$  if and only if the map  $x \rightarrow n \cdot x$  of  $\mathbb{Z}$  into  $G$  extends as a continuous homomorphism of  $\mathbb{Z}_p$  into  $G$ .

(ii) If  $G$  is totally disconnected then  $G_p$  is closed and is a  $\mathbb{Z}_p$ -module.

(iii) If  $G$  is  $p$ -primary, so is  $G^*$ . (For proofs and other details see [A] or [B]).

**Lemma 9** Let  $G \in \mathcal{C}$  i.e., both  $G$  and  $G^*$  are totally disconnected. Let  $K$  be a compact open subgroup of  $G$  and let  $G_p, K_p$  denote their  $p$ -primary components. Then  $K_p$  is a compact open subgroup of  $G_p$  and  $G \simeq \Pi G_p$  is the restricted direct product with respect to  $\{K_p\}$ . (For proof see [B]).

This leads us to introduce the class  $\mathcal{C}_p$ . Let  $\mathcal{C}_p$  consist of all  $G \in \mathcal{C}$  which satisfy the following conditions:

(i)  $G$  is  $p$ -primary,

(ii)  $\dim_{\mathbb{Q}_p} \text{hom}(G, \mathbb{Q}_p) < \infty$

(iii) The subgroup  $= \{x \in G: p \cdot x = 0\}$  of elements of order  $p$  is finite.

We then have.

## PROPOSITION 10

Let  $G \in \mathcal{C}_p$ . Then  $G$  is isomorphic to a group of the form  $G \simeq \mathbb{Q}_p^a \times \mathbb{Z}_p^b \times T_p^c \times B$ , where  $B$  is a finite  $p$ -group, where  $a, b, c$  are finite integers.

The proof will be carried out in a number of steps.

**Step 1.** Let  $\Gamma = \cap \{\ker \gamma: \gamma \in \text{hom}(G, \mathbb{Q}_p)\}$ . Then for some integer  $a, b, a + b = d = \dim \text{hom}(G, \mathbb{Q}_p)$ ,  $G/\Gamma \simeq \mathbb{Q}_p^a \times \mathbb{Z}_p^b$ .

*Proof.* Let  $\xi_1, \dots, \xi_d$  be a basis for  $\text{hom}(G, \mathbb{Q}_p)$ . Let  $\varphi: G \rightarrow \mathbb{Q}_p^d$ ,  $\varphi = (\xi_1, \dots, \xi_d)$ . Then the linear independence of  $\xi_j$ 's implies that  $\xi(G)$  spans  $\mathbb{Q}_p^d$ . Thus  $\xi(G)$  contains a basis of  $\mathbb{Q}_p^d$ , is a  $\mathbb{Z}_p$ -module and so  $\xi(G)$  is an open  $\mathbb{Z}_p$ -module or a lattice in  $\mathbb{Q}_p^d$ . From the known structure of such lattices (see [W2], Chapter 2), it follows that

$$\xi(G) = \sum_{j=1}^a \mathbb{Q}_p v_j + \sum_{a+1}^d \mathbb{Z}_p v_j \simeq \mathbb{Q}_p^a \times \mathbb{Z}_p^b.$$

**Step 2.** The subgroup  $\Gamma$  is discrete and  $\simeq T_p^c \times B$ , for some finite group  $B$ .

*Proof.* Using the earlier notation, let  $x_j \in G$  be such that  $\varphi(x_j) = v_j$ . Let  $M = \Sigma \mathbb{Z}_p x_j$ . Then  $M$  is a compact subgroup of  $G$ ,  $\Gamma \cap M = \langle 0 \rangle$  and  $\Gamma \cdot M = \varphi^{-1}(\Sigma \mathbb{Z}_p v_j)$  and so  $\Gamma \cdot M$  is open in  $G$ . Let  $\Gamma_0$  be a compact open subgroup of  $\Gamma$ . Then  $\Gamma_0 \cdot M$  is a compact

open subgroup of  $G$ . Then  $\Gamma_0^*$ —the dual group of  $\Gamma_0$  is a discrete  $p$ -primary group. Also  $\Gamma_0^*$  is reduced or has no divisible subgroup. For if it has, then being  $p$ -primary it has subgroups of the form  $T_p$  and  $\Gamma_0$  will then have quotients of the form  $\mathbb{Z}_p$  (for  $\mathbb{Z}_p$  is the character group of  $T_p$ ). But such a quotient gives rise to a homomorphism of  $\Gamma_0$  into  $\mathbb{Q}_p$ . This can be extended first to  $\Gamma_0 M$ , since  $\Gamma_0 \cap M = \langle 0 \rangle$ , and then to  $G$ , since  $\Gamma_0 M$  is open in  $G$ . But this contradicts that  $\Gamma_0 \subset \Gamma = \ker \text{hom}(G, \mathbb{Q}_p)$ . Thus  $\Gamma_0^*$  is reduced and so has a cyclic direct summand from a general theorem on abelian groups (see [K], page 21, Theorem 9).

Clearly then  $\Gamma_0$  also has a cyclic direct summand  $\Gamma_0 = C \cdot \Gamma_1$ , and the argument can be repeated for  $\Gamma_1$ . But the number of cyclic direct summands arising in this way has to be finite since  $\{x \in \Gamma_0 : p \cdot x = 0\}$  is finite. Thus  $\Gamma_0$  itself is finite. This implies that  $\Gamma$  is discrete, since  $\Gamma_0$  is open. Clearly  $\Gamma_d$ —the maximal divisible subgroup of  $\Gamma$ , is a direct summand and is  $\simeq T_p^c$ ,  $c$  being finite integer, since elements of order  $p$  in  $\Gamma$  are finite. Thus  $\Gamma = \Gamma_d \cdot B$ . The subgroup  $B$  is then finite by a similar argument.

Finally note that if  $L \in \mathcal{L}$  is  $p$ -primary and  $L_0 \subset L$  is closed subgroup then  $L_0$  is a direct summand either when  $L/L_0 \simeq \mathbb{Z}_p^b$  for some finite  $b$  or when  $L_0$  is discrete and  $\simeq T_p^b$ . Here the second part comes out of duality. Applying this to  $G$ , we get  $G \simeq T_p^c \times \mathbb{Z}_p^b \cdot H$ , where  $H$  is a closed subgroup, with the property that  $B \subset H$ ,  $H/B \simeq \mathbb{Q}_p^a$  and  $B$  is finite. If  $L$  is the annihilator of  $B$  in  $H^*$ , then  $L \simeq (\mathbb{Q}_p^a)^* \simeq \mathbb{Q}_p^a$  and  $H^*/L \simeq B^* \simeq B$ . Thus  $L$  is open. An open divisible subgroup is a direct summand and thus  $H^* \simeq L \cdot B^*$  or  $H \simeq \mathbb{Q}_p^a \cdot B$ .  $\square$

## COROLLARY 11

If  $G \in \mathcal{C}_p$ , so does  $G^*$ .

This is clear since character groups of the form  $\mathbb{Q}_p^a \times \mathbb{Z}_p^b \times T_p^c \times B_p$  are again of the same form. Also any group of the above form belongs to  $\mathcal{C}_p$ .

Putting together these results one can describe the structure of the class  $\mathcal{L}_0$  as follows. (See Theorem 1 in §1 for definition of the class  $\mathcal{L}_0$ ).

## PROPOSITION 12

If  $G \in \mathcal{L}_0$ , so does its dual  $G^*$ . Moreover there exist closed subgroups  $G_\infty, G_p$  such that  $G = G_\infty \cdot \Pi_p G_p$ —a restricted direct product with respect to  $\{K_p\}$ , where  $K_p$  is a compact open subgroup of  $G_p$ . Also the group  $G_\infty$  is isomorphic to a group of the form  $R^m \cdot T^n \cdot \mathbb{Z}^k$  and  $G_p \in \mathcal{C}_p$  for all  $p$ .

## 3. Symplectic spaces $(X, \omega)$

Let  $\omega(\cdot, \cdot)$  be a continuous alternating bicharacter of  $X$ ,  $X \in \mathcal{L}$ . The alternating property viz:  $\omega(x, x) = 1$  for all  $x \in X$  implies that  $\omega(x, y) = \omega(y, x)^{-1}$ , for all  $x, y \in X$ . If  $\omega_x: y \rightarrow \omega(x, y)$  is the character of  $X$  defined by  $x$ , then nondegeneracy is equivalent to the statement that  $x \rightarrow \omega_x$  is an isomorphism of  $X$  with its dual  $X^*$ . Thus, in what follows we consider symplectic spaces  $(X, \omega)$ ,  $X \in \mathcal{L}$ ,  $\omega$  nondegenerate, and identify

for all  $x \in Z$  and (ii)  $X = (Z_*) \cdot U$  direct sum. Then  $X_1 = Z \cdot U$  is closed; and if  $X_2 = (X_1)_*$ , then  $X = X_1 \cdot X_2$  direct sum. Note this gives a decomposition of  $X$  as orthogonal direct sum of symplectic spaces  $(X_j, \omega)$ .

*Proof.* Since  $\omega$  is nondegenerate,  $X/Z_*$  is isomorphic to the character group of  $Z$ . In particular the map  $x, u \rightarrow \omega(x, u)$  is a dual pairing of  $Z$  and  $U$ . Let  $H = U_* \cap Z_*$ . Then  $H$  is closed. Let  $x_0 \in Z_*$  be arbitrary. Consider the character  $\omega_{x_0}|U$ . Then there exists an element  $z_0 \in Z$ , such that  $\omega_{x_0} = \omega_{z_0}$  on  $U$ . Thus  $x_0 = z_0 \cdot h$ , for some  $h \in H$ . Since we are working within the class of second countable groups, it follows that  $Z_* = Z \cdot H$  direct sum and thus  $X = Z \cdot H \cdot U$  direct sum. This implies that  $X_1 = Z \cdot U$  is closed, and if we write  $X_2 = H$ , then  $X = X_1 \cdot X_2$  is an orthogonal symplectic decomposition.

#### PROPOSITION 14.

Let  $(X, \omega)$  be symplectic. Assume that the maximal compact connected subgroup of  $X$  is a torus. Then there exists an orthogonal symplectic decomposition  $X = X_1 \cdot X_2 \cdot X_3$  (i.e.,  $\omega(X_i, X_j) = 1$  for  $i \neq j$  and  $(X_j, \omega)$  are nondegenerate) such that  $X_1 \simeq R^{2n}$ ,  $X_2 \simeq T^k \times \mathbb{Z}^k$  and  $X_3$  is totally disconnected or  $X_3 \in \mathcal{C}$ .

*Proof.* Let  $K$  be the maximal compact connected subgroup of  $X$ . Then  $K = ({}^0X \cap X^0)$  and  $K_* = ({}^0X \cdot X^0)$ . Also  $X/K_*$  is isomorphic to the character group of  $K$  and so  $\simeq \mathbb{Z}^k$ . Thus  $K_*$  is direct summand. Or there exists a closed subgroup  $D$ ,  $\simeq \mathbb{Z}^k$  such that  $X \simeq K_* \cdot D$ . Thus by Lemma 13,  $X = X_1 \cdot Y$  with  $X_1 = K \cdot D$  and  $Y = (X_1)_*$ . Since  $(Y, \omega)$  is nondegenerate and  $Y$  has no compact connected subgroups, it follows that  $Y = Y^0({}^0Y)$  direct sum,  $Y^0 \simeq R^m$  and  ${}^0Y$  is totally disconnected. Since the character group of  ${}^0Y$  is also totally disconnected it follows that  $\omega(Y^0, {}^0Y) = 1$ . Thus if  $X_2 = Y^0$  and  $X_3 = {}^0Y$ ,  $X = X_1 \cdot X_2 \cdot X_3$  is the required decomposition.

#### PROPOSITION 15.

Let  $(X, \omega)$  be nondegenerate. Assume that  $X$  is totally disconnected. Let  $X_p$  be the  $p$ -primary component of  $X$ . Then  $\omega(X_p, X_q) = 1$  where  $p \neq q$ , and  $(X_p, \omega)$  is nondegenerate. If  $K$  is any compact open subgroup of  $X$ , such that  $\omega(K, K) = 1$ , then  $X = \prod_p X_p$ —a restricted direct product with respect to  $\{K_p\}$  is also an orthogonal symplectic decomposition of  $X$ .

*Proof.* Note that if  $\chi$  is a continuous character of a  $p$ -primary group  $G$ , then  $\chi(G) \subset T_p$ . Thus if  $x \in X_p$  and  $y \in X_q$ ,  $p \neq q$ , then  $\omega(x, y) \in T_p \cap T_q = \langle 1 \rangle$ . Thus  $\omega(X_p, X_q) = 1$ . Next let  $U$  be a neighborhood of the identity in  $T$ , such that  $U$  contains no subgroups other than the trivial one. Consider the open subset  $W = \{(x, y) \in X \times X : \omega(x, y) \in U\}$ . Since compact open subgroups form a basis of neighborhoods of the identity in  $X$ , there exist a compact open subgroup  $K$ , such that  $K \times K \subset W$ . Thus  $\omega(K, K) = 1$ . The rest follows easily (see Lemma 9).

#### PROPOSITION 16.

Let  $(X, \omega)$  be nondegenerate with  $X \in \mathcal{C}_p$ . Then there exists an orthogonal symplectic decomposition  $X = X_1 \cdot X_2 \cdot X_3$  (direct sum) such that  $X_1 \simeq T_p^n \times \mathbb{Z}_p^n$ ,  $X_2 = \mathbb{Q}_p^m$  and  $X_3$  is finite.

*Proof.* Since  $X \in \mathcal{C}_p$ , let  $X = H_1 \cdot H_2 \cdot H_3 \cdot H_4$ , where  $H_1 \simeq \mathbb{Q}_p^m$ ,  $H_2 \cong T_p^n$ ,  $H_3 \simeq \mathbb{Z}_p^k$ ,  $H_4$  finite. Then  $H_2 H_4$  is the closed torsion subgroup  $X_0$  of  $X$  and  $H_2$  is the divisible part of  $X_0$ . Since  $X \simeq X^*$  and the divisible part of the torsion subgroup of  $X^* \simeq H_3^* \simeq T_p^k$ . Thus  $T_p^n \simeq T_p^k$  and so  $n = k$ . Next since  $H_1 H_2$  is divisible, while  $H_2 H_4$  is torsion, it follows that  $\omega(H_1 H_2, H_2 H_4) = 1$ . Thus  $(H_2)_* = H_1 \cdot H_2 \cdot D \cdot H_4$ , for some closed subgroup  $D$  of  $H_3$ . So  $X/(H_2)_* \simeq H_3/D$  and on the other hand  $\simeq H_2^* \simeq \mathbb{Z}_p^n \simeq H_3$ . Thus  $H_3 \simeq H_3/D$ , forces  $D = \langle 1 \rangle$ , or  $(H_2)_* = H_1 H_2 H_4$ . If you now use Lemma 13, it follows that  $X = X_1 \cdot Y$  is an orthogonal symplectic decomposition with  $X_1 = H_2 H_3$  and  $Y = (X_1)_*$ . Clearly  $Y \simeq X/X_1 \simeq H_1 H_4$ . Thus  $Y_d$ —the maximal divisible subgroup of  $Y \simeq H_1 \simeq \mathbb{Q}_p^m$  and the torsion subgroup  $Y_0$  of  $Y$  is  $\simeq H_4$ . Clearly  $\omega(Y_d, Y_0) = 1$ , and  $Y = Y_d \cdot Y_0$ . Thus if we take  $X_2 = Y_d, X_3 = Y_0$ , then  $X = X_1 \cdot X_2 \cdot X_3$  is an orthogonal symplectic decomposition with the required properties.

As a final step in the proof of Theorem 1 we need the following.

*Lemma 17.* Let  $(X, \omega)$  be nondegenerate. In each of the following cases (i)  $X \simeq R^m$ , (ii)  $X \simeq T^n \times \mathbb{Z}^n$ , (iii)  $X \simeq \mathbb{Q}_p^m$ , (iv)  $X \simeq T_p^n \times \mathbb{Z}_p^n$  and (v)  $X$  is finite,  $(X, \omega)$  admits a polarization. Moreover the polarizing groups  $G, G^*$  are of some type as  $X$ .

*Proof.* In cases (i) and (iii),  $\omega(x, y) = \chi(\Omega(x, y))$  for some nondegenerate symplectic bilinear form on  $X (= R^m$  or  $\mathbb{Q}_p^m)$ , where  $\chi$  is a non-trivial character on  $R$  or  $\mathbb{Q}_p$ . Existence of symplectic bases for  $\Omega$  gives polarizations for  $X$ . The case (v) is a classical result of Frobenius. The cases (ii) and (iv) are proved similarly. We sketch the argument for (ii). Note  $X^0 \simeq T^n$ . Choose a closed subgroup  $S_1 \subset X^0, S_1 \simeq T$ . Then  $X/(S_1)_* \simeq \mathbb{Z}$ . Thus  $X = (S_1)_* \cdot D_1$ , for some closed subgroup  $D_1 \simeq \mathbb{Z}$ . By Lemma 13,  $X = X_1 \cdot X_2$  is an orthogonal symplectic decomposition,  $X_1 = S_1 \cdot D_1$ , the subgroups  $S_1, D_1$  defining a polarization for  $X_1$ . Note  $X_2$  is again of the same type and so induction works. The case (iv) is handled similarly.

*Remark.* Although the class of groups  $\mathcal{L}_0$  includes most of the examples arising in applications, it is not known whether, for a general  $X \in \mathcal{L}$ , a nondegenerate symplectic structure always admits a polarization.

## Acknowledgements

This work was done while the author was visiting the Tata Institute in Fall 92 and the author would like to thank Professor M S Raghunathan for his hospitality and for many stimulating conversations.

## References

- [A] Armacost D L, *The structure of locally compact abelian groups* (Marcel Dekker, Inc) (1981)
- [B] Braconnier J, Sur les groupes topologiques localement compacts, *J. Math. Pure. Appl.*, N. S. 27 1–85 (1948)
- [HR] Hewitt E and Ross K A, *Abstract harmonic analysis*, (Springer-Verlag) Vol. I (1963)
- [K] Kaplansky I, *Infinite abelian groups* (University of Michigan Press, Ann Arbor) (1969)
- [M] Mumford D, (with Madhav Nori and Peter Norman), *Tata Lectures on Theta III*, Birkhauser (1992)
- [W1] Weil A, Sur certains groupes des operateurs unitaires, *Acta Math.* 111 (1964) 143–211
- [W2] Weil A, *Basic number theory* (Springer-Verlag) (1974)



# Modular equations and Ramanujan's Chapter 16, Entry 29

GEORGE E ANDREWS

Department of Mathematics, Pennsylvania State University, 410 McAllister Building,  
University Park, PA 16802, USA

Dedicated to the memory of my friend, Professor K G Ramanathan

**Abstract.** In this paper we illustrate how some of the classical modular equations can be proved by using only Ramanujan's summation (see (1.1)) and dispensing completely with the Schröter-type methods.

**Keywords.** Modular equations; Rogers-Ramanujan functions.

## 1. Introduction

I first met K G Ramanathan because of our mutual interest in Ramanujan. He came to Penn State in February 1982 at my invitation to give our colloquium. I had been through a particularly trying week and was exhausted to say the least. Ramanathan presented a beautiful lecture explaining and extending work from Ramanujan's Lost Notebook [9]–[13]. I remember few mathematics talks as fondly as I remember that one. The beauty of the work truly revived my spirits.

General interest in Ramanujan's work has been intense in recent years due in no small part to the magnificent edited versions of Ramanujan's Notebooks [2], [3], [4] carefully prepared by Bruce Berndt.

This paper will be devoted to further considerations of modular equations, especially those of degrees 3 and 5. Berndt [4; pp. 6–7] and Hardy [7; Ch. 12] discuss several approaches to modular equations. Succinctly stated they are: (1) the Legendre-Jacobi method using differential equations for elliptic functions [7; §§ 12.4–12.7]; (2) Schröter's method requiring ingenious rearrangements of double theta series [4; p. 73]; (3) the theory of modular forms [4; p. 7], and (4) Ramanujan's method.

Both Hardy and Berndt are uncertain about Ramanujan's method for the excellent reason that he never revealed it. He merely stated his discoveries without proof, and as Berndt puts it "... found more modular equations than all of his predecessors put together."

To prove Ramanujan's formulas both Hardy [7; Ch. 12] and Berndt [4] mix Schröter's method, algebraic manipulation of series and products (what Hardy [7; pp. 220–221] calls "trivial" relations), and Ramanujan's  ${}_1\psi_1$ -summation [4; p. 32, Entry 17] rewritten as

$$\sum_{n=-\infty}^{\infty} \frac{(a)_n t^n}{(b)_n} = \frac{(b/a, at, q/(at), q; q)_{\infty}}{(q/a, b/(at), b, t; q)_{\infty}}, \quad (1.1)$$

where

$$(A)_n = (A; q)_n = \prod_{j=0}^{\infty} \frac{(1 - Aq^j)}{(1 - Aq^{j+n})} \quad (1.2)$$

( $= (1 - A)(1 - Aq) \dots (1 - Aq^{n-1})$  when  $n$  is a positive integer),

$$(A)_{\infty} = (A; q)_{\infty} = \lim_{n \rightarrow \infty} (A)_n, \quad (1.3)$$

and

$$(A_1 A_2, \dots, A_r; q)_{\infty} = (A_1)_{\infty} (A_2)_{\infty} \dots (A_r)_{\infty}. \quad (1.4)$$

Actually the only instance of (1.1) required in this regard is the case  $b = aq$ :

$$\begin{aligned} S(a, t, q) &:= \frac{(at, q/(at), q, q; q)_{\infty}}{(a, q/a, t, q/t; q)_{\infty}} \\ &= \sum_{n=-\infty}^{\infty} \frac{t^n}{1 - aq^n} \end{aligned} \quad (1.5)$$

(where for convergence we require  $|q| < |t| < 1$ ).

Our object in this paper is to show that the Schröter methods may be entirely dispensed with at least for the standard forms of the modular equations of degrees 3 and 5. I am certainly not suggesting that Ramanujan did not know Schröter's method. However I would stress that the methods given here at each stage suggest very simple combinations of functions (see especially the proof of Entry 29 in §3 below) which translate into complicated and surprising modular equations.

This approach appears to fit in nicely with the first 22 entries considered by Berndt in [5]. This latter paper is devoted to the results on theta-functions and modular equations found in the 100 pages of unorganized material at the end of Ramanujan's second notebook and in the 33 pages of unorganized material comprising the third notebook. The reader's attention is also directed to the recent work of L.-C. Shen [14] who also uses (1.1) to derive Lambert series identities related to modular equations of degrees 3.

## 2. Background

In order to maintain the Ramanujan spirit, we follow Berndt's lead and work with [4; p. 34, p. 36]

$$\varphi(q) := \sum_{n=-\infty}^{\infty} q^{n^2} \quad (2.1)$$

$$\psi(q) := \sum_{n=-\infty}^{\infty} q^{n(n+1)/2} = \sum_{n=-\infty}^{\infty} q^{n(2n+1)}. \quad (2.2)$$

$$f(a, b) := \sum_{n=-\infty}^{\infty} a^{n(n+1)/2} b^{n(n-1)/2}. \quad (2.3)$$

If in (1.1) we replace  $t$  by  $t/a$ , set  $b = 0$  and then let  $a \rightarrow \infty$ , we obtain Jacobi's triple



$$\sum_{n=-\infty}^{\infty} (-1)^n q^{n(n-1)/2} t^n = (t, q/t, q; q)_{\infty}, \quad (2.4)$$

which is equivalent to (with  $t = -a, q = ab$ )

$$f(a, b) = (-a, -b, ab; ab)_{\infty} \quad (2.5)$$

Still following Berndt's account of the basics [4; p. 36, Entry 22(i), (ii); p. 37 eq. (22.4); p. 40, Entry 25, (i)–(iv)]

$$\varphi(q) = \frac{(-q, q^2; q^2)_{\infty}}{(q, -q^2; q^2)_{\infty}} \quad (2.6)$$

$$\varphi(-q) = \frac{(q)_{\infty}}{(-q; q)_{\infty}} \quad (2.7)$$

$$\psi(q) = \frac{(q^2, q^2)_{\infty}}{(q; q^2)_{\infty}} \quad (2.8)$$

$$\varphi(q) = \varphi(q^4) + 2q\psi(q^8) \quad (2.9)$$

$$\varphi(q)\varphi(-q) = \varphi^2(-q^2) \quad (2.10)$$

$$\psi(q)\psi(-q) = \psi(q^2)\varphi(-q^2) \quad (2.11)$$

$$\varphi(q)\psi(q^2) = \psi^2(q). \quad (2.12)$$

An examination of Berndt's clear presentation [4; pp. 36–40] shows the direct derivation of each expression either from (2.5) or from the algebraic manipulation of infinite series and products.

### 3. Entry 29 of Chapter 16

If  $ab = cd$ , then

$$\begin{aligned} f(a, b)f(c, d) + f(-a, -b)f(-c, -d) \\ = 2f(ac, bd)f(ad, bc), \end{aligned} \quad (3.1)$$

and

$$\begin{aligned} f(a, b)f(c, d) - f(-a, -b)f(-c, -d) \\ = 2af\left(\frac{b}{c}, \frac{c}{b}abcd\right)f\left(\frac{b}{d}, \frac{d}{b}abcd\right). \end{aligned} \quad (3.2)$$

Berndt's proof of these [4; p. 45] is a nice application of the rearrangement of double series. We shall show that these results follow from Ramanujan's  ${}_1\psi_1$ -summation as stated in (1.5) above.

Obviously

$$S(A, t, q) \pm S(A, -t, q) = \sum_{n=-\infty}^{\infty} \frac{t^n(1 \pm (-1)^n)}{1 - Aq^n}; \quad (3.3)$$

So

$$S(A, t, q) + S(A, -t, q) = 2S(A, t^2, q^2), \quad (3.4)$$

$$S(A, t, q) - S(A, -t, q) = 2t S(Aq, t^2, q^2). \quad (3.5)$$

Equation (3.4) simplifies to (3.1) with  $t = c$ ,  $A = -a/c$ ,  $q = ab$ , and (3.5) reduces to (3.2) under the same substitution.

For our purposes, Entry 29 given by (3.4) and (3.5) is the most useful. We shall concentrate on the two following specializations. Let

$$\begin{aligned} F_{k, \cdot}(q) &= \frac{(-q^{2l+1}, -q^{2k-2l-1}, q^{2k}, q^{2k}; q^{2k})_{\infty}}{(q^{2l+1}, q^{2k-2l-1}, -q^{2k}, -q^{2k}; q^{2k})_{\infty}} \\ &= \varphi^2(-q^{2k}) \frac{(-q^{2l+1}, -q^{2k-2l-1}, q^{2k})_{\infty}}{(q^{2l+1}, q^{2k-2l-1}; q^{2k})_{\infty}} \\ &= 2S(-1, q^{2l+1}, q^{2k}) \\ &= \sum_{n=-\infty}^{\infty} \frac{q^{n(2l+1)}}{1 + q^{2kn}} \end{aligned} \quad (3.6)$$

and

$$\begin{aligned} G_{k, l}(q) &= \frac{(-q^{2l+k+1}, -q^{k-2l-1}, q^{2k}, q^{2k}; q^{2k})_{\infty}}{(q^{2l+1}, q^{2k-2l-1}, -q^k, -q^k; q^{2k})_{\infty}} \\ &= \psi^2(-q^k) \frac{(-q^{2l+k+1}, -q^{k-2l-1}; q^{2k})_{\infty}}{(q^{2l+1}, q^{2k-2l-1}; q^{2k})_{\infty}} \\ &= S(-q^k, q^{2l+1}, q^{2k})_{\infty} \\ &= \sum_{n=-\infty}^{\infty} \frac{q^{n(2l+1)}}{1 + q^{k(2n+1)}}. \end{aligned} \quad (3.7)$$

So by (3.4)

$$F_{k, l}(q) + F_{k, l}(-q) = 2F_{k, l}(q^2), \quad (3.8)$$

and by (3.5)

$$F_{k, l}(q) - F_{k, l}(-q) = 4q^{2l+1} G_{k, l}(q^2). \quad (3.9)$$

#### 4. The modular equation of degree 1

This section is devoted to the case  $k = 1, l = 0$  of (3.6) and (3.7). Note that by (2.6)–(2.8):

$$F_{1, 0}(q) = \varphi^2(q), \quad (4.1)$$

and

$$G_{1, 0}(q) = 2\psi^2(q^2). \quad (4.2)$$

Hence by (3.8)

$$\varphi^2(q) + \varphi^2(-q) = 2\varphi^2(q^2), \quad (4.3)$$

and by (3.9)

$$\varphi^2(q) - \varphi^2(-q) = 8q\psi^2(q^4). \quad (4.4)$$

Equations (4.3) and (4.4) are in fact items (v) and (vi) in Entry 25 [4; p. 40]. Lastly, Berndt points out that multiplying them together yields

$$\varphi^4(q) - \varphi^4(-q) = 16q\psi^4(q^2). \quad (4.5)$$

This identity in the notation of classical elliptic function theory [6; p. 93 eq. (34.32)] is the identity

$$k^2 + k'^2 = 1, \quad (4.6)$$

an identity that could be called (but never is) the modular equation of degree 1.

## 5. The modular equation of degree 3

Now we consider (3.8) with  $k = 3$ ,  $l = 0$ :

$$F_{3,0}(q) = \frac{\varphi(-q^6)\varphi(-q^2)\varphi(-q^3)}{\varphi(-q)} \quad (5.1)$$

Thus after simplification, (3.8) reduces to

$$\varphi(q)\varphi(-q^3) + \varphi(-q)\varphi(q^3) - 2\varphi(-q^4)\varphi(-q^{12}) = 0. \quad (5.2)$$

This is one of many forms of the modular equation of degree 3. Legendre's standard form of the modular equation of degree 3 [4; p. 232] is

$$\varphi(q)\varphi(q^3) - \varphi(-q)\varphi(-q^3) - 4q\psi(q^2)\psi(q^6) = 0. \quad (5.3)$$

To see that (5.2) and (5.3) are equivalent, we apply (2.9) to (5.2):

$$\begin{aligned} 0 &= (\varphi(q^4) + 2q\psi(q^8))(\varphi(q^{12}) - 2q^3\psi(q^{24})) \\ &\quad + (\varphi(q^4) - 2q\psi(q^8))(\varphi(q^{12}) + 2q^3\psi(q^{24})) - 2\varphi(-q^4)\varphi(-q^{12}) \\ &= 2\varphi(q^4)\varphi(q^{12}) - 8\psi(q^8)\psi(q^{24}) - 2\varphi(-q^4)\varphi(-q^{12}). \end{aligned}$$

This reduces to (5.3) upon division by 2 and replacement of  $q^4$  by  $q$ .

## 6. Ramanujan's identities of modulus 5

In this section we consider the cases  $k = 5$ ,  $l = 0, 1$  of (3.8) and (3.9). We begin by recalling the Rogers-Ramanujan infinite products [4; pp. 77–78]:

$$g(q) = \frac{1}{(q, q^4; q^5)_\infty}, \quad (6.1)$$

$$(q^-, q^-, q^-)_\infty$$

Clearly from (6.1) and (6.2)

$$g(q)h(q) = \frac{(q^5; q^5)_\infty}{(q)_\infty}, \quad (6.3)$$

$$\frac{g(q)}{h(q^2)} = \frac{1}{(q, q^9; q^{10})_\infty}, \quad (6.4)$$

$$\frac{h(q)}{g(q^2)} = \frac{1}{(q^3, q^7; q^{10})_\infty}, \quad (6.5)$$

$$\frac{g(q)}{h(q^4)} = \frac{(-q^4, -q^6; q^{10})_\infty}{(q, q^9; q^{10})_\infty}, \quad (6.6)$$

and

$$\frac{h(q)}{g(q^4)} = \frac{(-q^2, -q^8; q^{10})_\infty}{(q^3, q^7; q^{10})_\infty}. \quad (6.7)$$

These formulae allow us to make the following identifications from (3.6) and (3.7):

$$F_{5,0}(q) = \frac{\varphi^2(-q^{10})g(q)}{g(-q)}, \quad (6.8)$$

$$F_{5,1}(q) = \frac{\varphi^2(-q^{10})h(q)}{h(-q)}, \quad (6.9)$$

$$G_{5,0}(q) = \frac{\psi^2(-q^5)g(q)}{h(q^4)}, \quad (6.10)$$

$$G_{5,1}(q) = \frac{\psi^2(-q^5)h(q)}{g(q^4)}. \quad (6.11)$$

by (6.3), (6.8) and (6.9)

$$F_{5,0}(q)F_{5,1}(q) = \frac{\varphi^3(-q^{10})\varphi(-q^5)\varphi(-q^2)}{\varphi(-q)}, \quad (6.12)$$

by (6.3), (6.8) and (6.11)

$$F_{5,0}(q)G_{5,1}(q) = \frac{\varphi^2(-q^{10})\psi^2(-q^5)(q^5; q^5)_\infty}{(q)_\infty g(-q)g(q^4)}, \quad (6.13)$$

and by (6.3), (6.9) and (6.10)

$$F_{5,1}(q)G_{5,0}(q) = \frac{\varphi^2(-q^{10})\psi^2(-q^5)(q^5; q^5)_\infty}{(q)_\infty h(-q)h(q^4)}. \quad (6.14)$$

n addition

$$\begin{aligned}
 \frac{h(q)}{g(q)} &= \frac{(q, q^4, q^6, q^9; q^{10})_{\infty}}{(q^2, q^3, q^7, q^8; q^{10})_{\infty}} \\
 &= \frac{g(q^2)}{h(q^2)^2 (q^{10}; q^{10})_{\infty}^2} \cdot \frac{(q, q^9, q^{10}, q^{10}; q^{10})_{\infty}}{(q^3, q^7, q^4, q^6; q^{10})_{\infty}} \\
 &= \frac{g(q^2)}{h(q^2)^2 (q^{10}; q^{10})_{\infty}^2} S(q^6, q^3, q^{10})_{\infty} \\
 &= \frac{g(q^2)}{h(q^2)^2 (q^{10}; q^{10})_{\infty}^2} \sum_{n=-\infty}^{\infty} \frac{q^{3n}}{1 - q^{10n+6}}; \tag{6.15}
 \end{aligned}$$

We conclude this list of specializations of  $S(A, t, q)$  with four which were listed earlier [1] in connection with the mock-theta conjectures:

$$\begin{aligned}
 \varphi(-q)g(q) &= \frac{S(-q, -q^2, q^5)}{(q^{10}; q^{10})_{\infty}} \\
 &= \frac{1}{(q^{10}; q^{10})_{\infty}} \sum_{n=-\infty}^{\infty} \frac{(-q^2)^n}{1 + q^{5n+1}} \\
 &\quad ([1; \text{p. 246, eq. (3.15)}]), \tag{6.16}
 \end{aligned}$$

$$\begin{aligned}
 \varphi(-q)h(q) &= \frac{S(-q, -q^3, q^5)}{(q^{10}; q^{10})_{\infty}} \\
 &= \frac{1}{(q^{10}; q^{10})_{\infty}} \sum_{n=-\infty}^{\infty} \frac{(-q^3)^n}{1 + q^{5n+1}} \\
 &= \frac{1}{(q^{10}; q^{10})_{\infty}} \sum_{n=-\infty}^{\infty} \frac{(-1)^{n-1} q^{2n+1}}{1 + q^{5n+4}} \\
 &\quad ([1; \text{p. 246, eq. (3.16)}]), \tag{6.17}
 \end{aligned}$$

$$\begin{aligned}
 \psi(q^2)g(q^4) &= \frac{1}{(q^{10}; q^{10})_{\infty}} \sum_{n=-\infty}^{\infty} \frac{q^{2n}}{1 - q^{20n+6}} \\
 &\quad ([1; \text{p. 247, eq. (3.18)}]), \tag{6.18}
 \end{aligned}$$

$$\begin{aligned}
 \psi(q^2)h(q^4) &= \frac{1}{(q^{10}; q^{10})_{\infty}} \sum_{n=-\infty}^{\infty} \frac{q^{2n}}{1 - q^{20n+14}} \\
 &\quad ([1; \text{p. 247, eq. (3.17)}]). \tag{6.19}
 \end{aligned}$$

We are now in a position to prove six formulas that are either due to Ramanujan directly deduced from Ramanujan's work by G N Watson.

$$\begin{aligned}
 \varphi(q)g(-q) - \varphi(-q)g(q) &= 2q\psi(q^2)h(q^4) \\
 &\quad ([16; \text{p. 289, eq. (6)}]), \tag{6.20}
 \end{aligned}$$

$$\begin{aligned}
 \varphi(q)h(-q) + \varphi(-q)h(q) &= 2\psi(q^2)g(q^4) \\
 &\quad ([16; \text{p. 289, eq. (7) corrected}], \tag{6.21}
 \end{aligned}$$

$$g(q)h(-q) + g(-q)h(q) = \frac{2\psi(q^2)}{(q^2; q^2)_\infty} = 2(-q^2; q^2)_\infty^2$$

([15; p. 60, eq. (5)]), (6.22)

$$g(q)h(-q) - g(-q)h(q) = \frac{2q\psi(q^{10})}{(q^2; q^2)_\infty}$$

([15; p. 60, eq. (6)]), (6.23)

$$g(q)g(q^4) + qh(q)h(q^4) = (-q; q^2)_\infty^2,$$

([15; p. 60, eq. (3)]), (6.24)

$$g(q)g(q^4) - qh(q)h(q^4) = \frac{\varphi(q^5)}{(q^2; q^2)_\infty}$$

([15; p. 60, eq. (4)]). (6.25)

By (6.16) and (6.19) we see that

$$\begin{aligned} & \varphi(q)g(-q) - \varphi(-q)g(q) \\ &= \frac{1}{(q^{10}; q^{10})_\infty} \sum_{n=-\infty}^{\infty} \left( \frac{(-q^2)^n}{1 - (-1)^n q^{5n+1}} - \frac{(-q^2)^n}{1 + q^{5n+1}} \right) \\ &= \frac{1}{(q^{10}; q^{10})_\infty} \sum_{n=-\infty}^{\infty} \left( \frac{q^{4n}}{1 - q^{10n+1}} - \frac{q^{4n}}{1 + q^{10n+1}} \right) \\ &= \frac{2}{(q^{10}; q^{10})_\infty} \sum_{n=-\infty}^{\infty} \frac{q^{14n+1}}{1 - q^{20n+2}} \\ &= \frac{2q}{(q^{10}; q^{10})_\infty} S(q^2, q^{14}, q^{20}) \\ &= \frac{2q}{(q^{10}; q^{10})_\infty} S(q^{14}, q^2, q^{20}) \\ &= \frac{2q}{(q^{10}; q^{10})_\infty} \sum_{n=-\infty}^{\infty} \frac{q^{2n}}{1 - q^{20n+14}} = 2q\psi(q^2)h(q^4), \end{aligned} \tag{6.26}$$

which is (6.20)

In exactly the same way (6.21) follows from (6.17) and (6.18). Now by (6.15), we see that

$$\begin{aligned} & g(q)h(-q) \pm g(-q)h(q) \\ &= g(q)g(-q) \left( \frac{h(-q)}{g(-q)} \pm \frac{h(q)}{g(q)} \right) \\ &= \frac{g(q)g(-q)g(q^2)}{h(q^2)^2 (q^{10}; q^{10})_\infty^2} \sum_{n=-\infty}^{\infty} \frac{q^{3n}((-1)^n \pm 1)}{1 - q^{10n+6}} \end{aligned} \tag{6.27}$$

Therefore the left-hand side of (6.22) is identical with

$$\begin{aligned}
 & \frac{2g(q)g(-q)g(q^2)}{h(q^2)^2(q^{10}; q^{10})_\infty^2} S(q^6, q^6, q^{20}) \\
 &= \frac{2g(q)g(-q)g(q^2)(q^{12}, q^8, q^{20}, q^{20}, q^{20})_\infty}{h(q^2)^2(q^{10}; q^{10})_\infty^2 (q^6, q^{14}, q^{20})_\infty^2} \\
 &= 2(-q^2; q^2)_\infty^2
 \end{aligned} \tag{6.28}$$

upon simplification using (6.1), (6.2) and putting all products to the modulus 20.

In exactly the same way, (6.27) implies that the left-hand side of (6.23) is identical with

$$\begin{aligned}
 & \frac{-2g(q)g(-q)g(q^2)q^3}{h(q^2)^2(q^{10}; q^{10})_\infty^2} S(q^{16}, q^6, q^{20}) \\
 &= \frac{2q\psi(q^{10})}{(q^2; q^2)_\infty}.
 \end{aligned} \tag{6.29}$$

To obtain (6.24) we observe that

$$\begin{aligned}
 & g(q)g(q^4) + qh(q)h(q^4) \\
 &= g(q) \left( \frac{\varphi(q)h(-q) + \varphi(-q)h(q)}{2\psi(q^2)} \right) \\
 & \quad + qh(q) \left( \frac{\varphi(q)g(-q) - \varphi(-q)g(q)}{2q\psi(q^2)} \right) \\
 &= \frac{\varphi(q)}{2\psi(q^2)} (g(q)h(-q) + h(q)g(-q)) \\
 &= \frac{\varphi(q)}{2\psi(q^2)} \frac{2\psi(q^2)}{(q^2; q^2)_\infty} = \frac{\varphi(q)}{(q^2; q^2)_\infty} = (-q; q^2)_\infty^2,
 \end{aligned} \tag{6.30}$$

as desired.

Finally

$$\begin{aligned}
 & \frac{(g(q)g(q^4) - qh(q)h(q^4))\psi^4(-q^5)}{h(q^4)g(q^4)} \\
 &= G_{5,0}(q) - qG_{5,1}(q) \\
 &= \frac{1}{2} \left( 2 \sum_{n=-\infty}^{\infty} \frac{q^n}{1 + q^{10n+5}} - 2 \sum_{n=-\infty}^{\infty} \frac{q^{3n+1}}{1 + q^{10n+5}} \right. \\
 & \quad \left. + \sum_{n=-\infty}^{\infty} \frac{q^{5n+2}}{1 + q^{10n+5}} \right) - \frac{1}{2} \sum_{n=-\infty}^{\infty} \frac{q^{5n+2}}{1 + q^{10n+5}} \\
 &= \frac{1}{2} \sum_{n=-\infty}^{\infty} \frac{q^n + q^{9n+4} - q^{3n+1} - q^{7n+3} + q^{5n+2}}{1 + q^{10n+5}} \\
 & \quad - \frac{q^2}{2} \sum_{n=-\infty}^{\infty} \frac{q^{5n}}{1 + q^{10n+5}}
 \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2} G_{1,0}(q) - \frac{1}{2} q^2 G_{1,0}(q^5) \\
&= \frac{1}{2} (\psi(q^2)^2 - q^2 \psi(q^{10})^2) \\
&= \frac{1}{2} \frac{(q^{10}; q^{10})_{\infty}^2}{g(q^2)h(q^2)}, \tag{6.31}
\end{aligned}$$

where to obtain the last expression we applied Entry 10 (v) [4; p. 262]. We note that Berndt's lovely proof of Entry 10 (v) is fully consistent with the object of this paper in that the only result used other than instances of rearrangements of  $S(A, t, q)$  is Entry 29 of Chapter 16 which we have shown to be again an application of Ramanujan's  ${}_1\psi_1$ -summation for  $S(A, t, q)$ . Equation (6.31) reduces to (6.25) upon product simplification.

## 7. The standard modular equation of degree 5

We shall consider the modular equation of degree 5 in the form given by Watson [16; p. 289]

$$\varphi(q)\varphi(-q^5) - \varphi(-q)\varphi(q^5) = 4q(q^4; q^4)_{\infty}(q^{20}; q^{20})_{\infty} \tag{7.1}$$

The proof goes rapidly using the results of § 6.

$$\begin{aligned}
&\varphi(q)\varphi(-q^5) - \varphi(-q)\varphi(q^5) \\
&= \varphi(q)\varphi(-q) \left( \frac{\varphi(-q^5)}{\varphi(-q)} - \frac{\varphi(q^5)}{\varphi(q)} \right) \\
&= \frac{\varphi(-q^2)\varphi(-q)}{\varphi^3(-q^{10})} (F_{5,0}(q)F_{5,1}(q) - F_{5,0}(-q)F_{5,1}(-q)) \\
&= \frac{\varphi(-q^2)\varphi(-q)}{2\varphi^3(-q^{10})} \{ (F_{5,0}(q) + F_{5,0}(-q))(F_{5,1}(q) - F_{5,1}(-q)) \\
&\quad + (F_{5,0}(q) - F_{5,0}(-q))(F_{5,1}(q) + F_{5,1}(-q)) \} \\
&= \frac{\varphi(-q^2)\varphi(-q)}{2\varphi^3(-q^{10})} \{ 8q^3 G_{5,0}(q^2)F_{5,1}(q^2) + 8q G_{5,0}(q^2)F_{5,1}(q^2) \} \\
&\hspace{25em} \text{(by (3.8) and (3.9))} \\
&= \frac{4q\varphi(-q^2)\varphi(-q)}{\varphi^3(-q^{10})} \left\{ \frac{\varphi^2(-q^{20})\psi^2(-q^{10})(q^{10}; q^{10})_{\infty}}{(q^2; q^2)_{\infty}h(-q^2)h(q^8)} \right. \\
&\quad \left. + q^2 \frac{\varphi^2(-q^{20})\psi^2(-q^{10})(q^{10}; q^{10})_{\infty}}{(q^2; q^2)_{\infty}g(-q^2)g(q^8)} \right\} \tag{by (6.8)-(6.11)}
\end{aligned}$$



$$\begin{aligned}
&= \frac{4q\varphi(-q^2)\varphi(-q)\varphi^2(-q^{20})\psi^2(-q^{10})(q^{10};q^{10})_\infty}{\varphi^3(-q^{10})(q^2;q^2)_\infty h(-q^2)h(q^8)g(-q^2)g(q^8)} \\
&\quad (g(-q^2)g(q^8) - (-q^2)h(-q^2)h(q^8)) \\
&= \frac{4q\varphi(-q^2)\varphi(-q)\varphi^2(-q^{20})\psi^2(-q^{10})(q^{10};q^{10})_\infty}{\varphi^3(-q^{10})(q^2;q^2)_\infty h(-q^2)h(q^8)g(-q^2)g(q^8)} \cdot \frac{\varphi(-q^{10})}{(q^4;q^4)_\infty} \\
&\quad \text{(by (6.25))} \\
&= 4q(q^4;q^4)_\infty (q^{20};q^{20})_\infty. \tag{7.2}
\end{aligned}$$

The last reduction follows by writing each factor in the penultimate line as an infinite product on the modulus 20 and doing the relevant cancellation.

## 8. Conclusion

The approach described in this paper suggests that further examination of  $F_{k,l}(q)$  and  $G_{k,l}(q)$  is in order. Obviously higher order modular equations might well be derived from a study of  $F_{2k+1,0}(q)$ . In addition it is possible that Hickerson's grand treatment of the seventh order mock theta functions [7] may well be related to  $F_{7,l}(q)$  ( $l = 0, 1, 2$ ).

## Acknowledgement

This study was partially supported by a grant from National Science Foundation.

## References

- [1] Andrews G E and Garvan F G, Ramanujan's "Lost" Notebook VI: The mock theta conjectures, *Adv. Math.*, **73** (1989) pp. 242–255
- [2] Berndt B C, Ramanujan's Notebooks, Part I (Springer, Berlin and New York) (1985)
- [3] Berndt B C, Ramanujan's Notebooks, Part II (Springer, Berlin and New York) (1989)
- [4] Berndt B C, Ramanujan's Notebooks, Part III (Springer, Berlin and New York) (1991)
- [5] Berndt B C, Chapter 26, theta functions and modular equations, (to appear in Ramanujan's Notebooks, Part IV)
- [6] Fine N J, Basic hypergeometric series and applications, *Math. Surveys and Monographs*, No. 27, *Am. Math. Soc.*, Providence, (1988)
- [7] Hardy G H, Ramanujan, Cambridge University Press, 1940 (Reprinted: Chelsea, New York) (1959)
- [8] Hickerson D R, On the seventh order mock theta functions, *Invent. Math.*, **94** (1988), pp. 639–660
- [9] Ramanathan K G, Remarks on some series considered by Ramanujan, *J. Ind. Math. Soc.*, **46** (1982) pp. 107–136
- [10] Ramanathan K G, On Ramanujan's continued fraction, *Acta Arith.*, **43** (1984) pp. 209–226
- [11] Ramanathan K G, On the Rogers-Ramanujan continued fraction, *Proc. Indian Acad. Sci. (Math. Sci.)*, **93** (1984) pp. 67–77
- [12] Ramanathan K G, Ramanujan's continued fraction, *Indian J. Pure Appl. Math.*, **16**(7) (1985) pp. 695–724
- [13] Ramanujan S, The lost notebook and other unpublished papers, (Narosa, New Delhi) (1987)
- [14] Shen L-C, On the modular equations of degree 3, *Proc. Am. Math. Soc.*, (to appear)
- [15] Watson G N, Proof of certain identities in combinatory analysis, *J. Ind. Math. Soc.*, **20** (1934) pp. 57–69
- [16] Watson G N, Mock theta functions (2), *Proc. Lond. Math. Soc. (2)*, **42** (1937) pp. 274–304



# Gaussian quadrature in Ramanujan's Second Notebook

RICHARD ASKEY

Department of Mathematics, Van Vleck Hall, University of Wisconsin, 480 Lincoln Drive,  
Madison, WI 53706, USA

Dedicated to the memory of Professor K G Ramanathan

**Abstract.** Ramanujan's notebooks contain many approximations, usually without explanations. Some of his approximations to series are explained as quadrature formulas, usually of Gaussian type.

**Keywords.** Gaussian quadrature; series approximations; Ramanujan.

## 1. Introduction

K G Ramanathan was a gentle man who had a strong sense of duty. Part of his duty was the understanding of Ramanujan and his mathematics, and we can all feel pleased that he helped us understand some of the mathematics Ramanujan did. In light of his work on Ramanujan's work on modular functions, continued fractions, and hypergeometric and basic hypergeometric functions, it is appropriate to dedicate a paper to his memory which deals with material from Ramanujan's Notebooks. The particular questions below deal with orthogonal polynomials, although it is very unlikely Ramanujan knew this. He was just looking for nice approximations he could compute easily, and attractive explicit formulas.

Ramanujan's approximations to certain series which I can explain are:

$$\varphi(0) + \frac{x}{1!}\varphi(1) + \frac{x^2}{2!}\varphi(2) + \frac{x^3}{3!}\varphi(3) + \dots \quad (1.1)$$

$$= e^x \varphi(x) \text{ as the first approximation,} \quad (1.1a)$$

$$= e^x \left\{ \frac{\sqrt{1+4x}-1}{2\sqrt{1+4x}} \varphi\left(x + \frac{1+\sqrt{1+4x}}{2}\right) + \frac{\sqrt{1+4x}+1}{2\sqrt{1+4x}} \varphi\left(x + \frac{1-\sqrt{1+4x}}{2}\right) \right\} \quad (1.1b)$$

$$= e^x \left\{ \frac{2}{3} \varphi(x) + \frac{\sqrt{1+12x}-1}{6\sqrt{1+12x}} \varphi\left(x + \frac{1+\sqrt{1+12x}}{2}\right) + \frac{\sqrt{1+12x}+1}{6\sqrt{1+12x}} \varphi\left(x + \frac{1-\sqrt{1+12x}}{2}\right) \right\}. \quad (1.1c)$$

These appear on page 352 of [4]. The rest appear on page 349 in [4].

$$\frac{1}{n} \{ \varphi(x-n+1) + \varphi(x-n+3) + \cdots + \varphi(x+n-1) \} \quad (1.2)$$

$$= \varphi(x) \text{ as the first approximation,} \quad (12a)$$

$$= \frac{\varphi\left(x + \sqrt{\frac{n^2-1}{3}}\right) + \varphi\left(x - \sqrt{\frac{n^2-1}{3}}\right)}{2} \text{ as the second,} \quad (1.2b)$$

$$= \frac{5(n^2-1) \left\{ \varphi\left(x + \sqrt{\frac{3n^2-7}{5}}\right) + \varphi\left(x - \sqrt{\frac{3n^2-7}{5}}\right) \right\} + 8(n^2-4)\varphi(x)}{6(3n^2-7)} \quad (1.2c)$$

$$= \left( \frac{1}{4} - \frac{n^2-16}{6\beta} \right) \left\{ \varphi\left(x + \sqrt{\frac{\alpha+\beta}{7}}\right) + \varphi\left(x - \sqrt{\frac{\alpha+\beta}{7}}\right) \right\} \\ + \left( \frac{1}{4} + \frac{n^2-16}{6\beta} \right) \left\{ \varphi\left(x + \sqrt{\frac{\alpha-\beta}{7}}\right) + \varphi\left(x - \sqrt{\frac{\alpha-\beta}{7}}\right) \right\} \quad (1.2d)$$

where  $\alpha = 3n^2 - 13$  and  $\beta = \sqrt{\frac{4}{5}(6n^4 - 45n^2 + 164)}$ .

He also included some examples

$$u_1 + u_2 + \cdots + u_{13} = \frac{13}{25}(7u_2 + 11u_7 + 7u_{12}) \quad (1.3)$$

$$u_1 + u_2 + \cdots + u_{22} \\ = \frac{11}{289}(161u_3 + 256u_{11\frac{1}{2}} + 161u_{20}) \quad (1.4)$$

$$u_1 + u_2 + \cdots + u_7 = \frac{7}{2}(u_2 + u_6) \quad (1.5)$$

$$u_1 + u_2 + \cdots + u_{26} = 13(u_6 + u_{21}) \quad (1.6)$$

$$\varphi(1) + \varphi(2) + \cdots + \varphi(21) \\ = \frac{7}{958} \left[ 506 \{ \varphi(2) + \varphi(20) \} + 931 \left\{ \varphi(1) + \varphi\left(11 + 2\sqrt{\frac{22}{7}}\right) \right. \right. \\ \left. \left. + \varphi\left(11 - 2\sqrt{\frac{22}{7}}\right) \right\} \right]. \quad (1.7)$$

## 2. Gaussian quadrature

Let  $f(t)$  be a continuous function on an interval  $[a, b]$ , and  $d\alpha(t)$  a non-negative

$$\int_a^b f(t) d\alpha(t) \quad (2.1)$$

by a finite sum which is exact for all polynomials of as high a degree as possible. When  $a < t_1 < \dots < t_k < b$ , set

$$w_k(t) = \prod_{i=1}^k (t - t_i) \quad (2.2)$$

and

$$w_{j,k}(t) = \frac{w_k(t)}{w'_k(t_j)(t - t_j)}. \quad (2.3)$$

Then

$$L_k^f(t) = \sum_{j=1}^n f(t_j) w_{j,k}(t) \quad (2.4)$$

is a polynomial of degree at most  $(k-1)$ , and

$$L_k^f(t_j) = f(t_j), \quad j = 1, 2, \dots, k. \quad (2.5)$$

When  $f(t)$  is a polynomial of degree  $(k-1)$ , then

$$\int_a^b f(t) d\alpha(t) = \sum_{j=1}^k f(t_j) \int_a^b w_{j,k}(t) d\alpha(t) \quad (2.6)$$

since

$$f(t) = L_k^f(t) \quad (2.7)$$

for all  $t$ . The degree  $(k-1)$  can be increased by an appropriate choice of the points  $t_j$ . If  $f(t)$  is a polynomial of degree  $(2k-1)$ , then

$$f(t) - L_k^f(t) = w_k(t) r_{k-1}(t) \quad (2.8)$$

with  $r_{k-1}(t)$  a polynomial of degree  $(k-1)$ . If  $w_k(t)$  is orthogonal to all polynomials of degree less than  $k$ , using the measure  $d\alpha(x)$  to define the inner product, then

$$\int_a^b f(t) d\alpha(t) - \int_a^b L_k^f(t) d\alpha(t) = \int_a^b w_k(t) r_{k-1}(t) d\alpha(t) = 0. \quad (2.9)$$

If

$$\lambda_j = \lambda_{j,k} = \int_a^b w_{j,k}(t) d\alpha(t) \quad (2.10)$$

then the Gaussian quadrature approximation to (2.1) is

$$\sum_{j=1}^k \lambda_j f(t_j). \quad (2.11)$$

There are other expressions for  $\lambda_j$  defined in (2.10). See Theorem 3.42 in Szegő [6] for three other expressions. Two of these expressions show immediately that  $\lambda_j > 0$ .

### 3. Ramanujan's claims

To obtain Ramanujan's claims, it is first necessary to identify the measure  $d\alpha(t)$ , and then to locate the points where the interpolation is done. Finally, the weighting coefficients must be obtained.

In example (1.1), the measure Ramanujan is using is obtained by multiplying both sides of the identities by  $e^{-x}$ . The measure is the Poisson distribution

$$\frac{e^{-x} x^j}{j!}, \quad j = 0, 1, \dots, \quad (3.1)$$

so  $a = 0$ ,  $b = \infty$  and the measure is the sum of infinitely many multiples of a shifted delta function.

The orthogonal polynomials for this measure are called Charlier polynomials, and they can be given as a hypergeometric series. The polynomials in (1.2) are also hypergeometric functions, so we recall the definition of a generalized hypergeometric series. This is a series whose term ratio is a rational function.

If the shifted factorial is defined by

$$\begin{aligned} (a)_n &= a(a+1)\cdots(a+n-1), & n = 1, 2, \dots, \\ 1, & & n = 0, \end{aligned} \quad (3.2)$$

then the hypergeometric series is

$${}_pF_q\left(\begin{matrix} a_1, \dots, a_p \\ b_1, \dots, b_q \end{matrix}; y\right) = \sum_{n=0}^{\infty} \frac{(a_1)_n \cdots (a_p)_n}{(b_1)_n \cdots (b_q)_n n!} y^n. \quad (3.3)$$

This usually requires that  $p \leq q + 1$ , for the series diverges if  $p > q + 1$  and it does not terminate. The Charlier polynomials are defined by

$$C_n(j; x) = {}_2F_0\left(\begin{matrix} -n, -j \\ - \end{matrix}; -\frac{1}{x}\right), \quad n = 0, 1, \dots \quad (3.4)$$

Since this series terminates, divergence is not a problem. See [3].

To obtain the zero used in (1.1a)

$$C_1(j; x) = 1 - \frac{j}{x}$$

so it vanishes when  $j = x$ . It is easy to check that formula (1.1a) is exact when

$$\varphi(x) = ax + b,$$

for it is clearly exact when  $a = 0$ , and when  $b = 0$  the calculation is routine.

A second interpretation of this approximation was given by Ramanujan in Chapter 3 of [4]. See Entry 10 in Berndt's version [1].

$$C_2(j; x) = 1 - \frac{2j}{x} + \frac{j(j-1)}{x^2}$$

as Ramanujan claimed.

The fact that the coefficients given by Ramanujan can be checked in two ways. Either one of the standard formulas can be used to derive them, or Ramanujan's formula can be used to check that there is equality for cubic polynomials.

A similar argument can be tried for the third approximation

$$C_3(j; x) = 1 - \frac{3j}{x} + \frac{3j(j-1)}{x^2} - \frac{j(j-1)(j-2)}{x^3}. \quad (3.4)$$

Ramanujan has the interpolation points at

$$j = x, x + \frac{1 + \sqrt{1 + 12x}}{2} \text{ and } x + \frac{1 - \sqrt{1 + 12x}}{2}.$$

$C_3(j; x)$  does not vanish at any of these points, so this is not a Gaussian quadrature formula. A Gaussian formula exists, but the zeros of (3.4) cannot be found as a simple expression, so Ramanujan did something else here. He took the value  $j = x$  as one interpolation point, which is reasonable for it is the expected value of the Poisson distribution. The remaining two points were chosen so the formula is exact for polynomials of maximal degree, which is four. This is most easily checked by showing there is equality for polynomials of degree 4. This is a tedious calculation which will not be given here.

The remaining formulas are all Gaussian quadrature formulas. In all the cases Ramanujan is using a uniform distribution on an equally spaced set of points. The usual notation for this takes the points at  $j = 0, 1, \dots, N$ . The polynomials orthogonal with respect to this distribution were found by Tchebychef [7]. They are given by a hypergeometric series as

$$Q_k(j; N) = {}_3F_2 \left( \begin{matrix} -k, k+1, -j \\ 1, -N \end{matrix}; 1 \right) \quad (3.5)$$

where  $j, k = 0, 1, \dots, N$ . This is the usual method of taking care of the zero which will appear in the denominator because  $(-N)_n = 0$  when  $n = N + 1, \dots$ . The two factors  $(-j)_n$  and  $(-k)_n$  both vanish when  $(-N)_n$  vanishes, and so the series continues to vanish when one zero in the numerator cancels a zero in the denominator. However, Ramanujan does not restrict his interpolation points to the integers, so we will define

$$Q_k(j; N) = \sum_{n=0}^k \frac{(-k)_n (k+1)_n (-j)_n}{(1)_n (-N)_n n!}$$

when  $k = 0, 1, \dots, N$ , but  $j$  is now allowed to be real or complex.

Again, we need to check the zeros of this function. In the previous case Ramanujan discovered Gaussian quadrature formulas when the polynomial was of degree 2, but for a cubic he did something else. In the present case, he goes up to degree 4, which

is possible because the polynomials are even or odd about the midpoint of the interval of orthogonality depending on the parity of the degree. Thus, such polynomials of degree 3 and 4 can be solved by taking square roots.

The examples (1.3)–(1.7) are instances of the general formulas in (1.2), after the step size has been changed. Ramanujan took step size 2 in (1.2) to avoid fractions in the first expression, but went back to the more usual step size of 1 in the examples (1.3)–(1.7).

#### 4. Comments

The polynomials in (3.5) are special cases of more general orthogonal polynomials. These polynomials.

$$Q_n(x; \alpha, \beta, N) = \sum_{k=0}^n \frac{(-n)_k (n + \alpha + \beta + 1)_k (-x)_k}{(\alpha + 1)_k (-N)_k k!}, \quad (4.1)$$

to revert to the more standard use of letters, are called Hahn polynomials. They are orthogonal on  $x = 0, 1, \dots, N$  with respect to the distribution

$$\binom{x + \alpha}{x} \binom{N - x + \beta}{N - x}, \quad x = 0, 1, \dots, N. \quad (4.2)$$

See [3]. The functions in (4.1) are multiples of what are called  $3-j$  symbols in quantum angular momentum theory. The location of integer zeros of  $3-j$  symbols is of some interest in mathematical physics. See [5] for some recent work.

In the introduction, I wrote that it is very unlikely Ramanujan was aware of the orthogonal polynomials which determine Gaussian quadrature. That should not be a surprise, for Gauss did not use orthogonality explicitly when he discovered Gaussian quadrature. Jacobi was the first to make this connection. There are a number of instances when Ramanujan seems to come close to orthogonal polynomials. This is especially true in some of his continued fractions, for the three term recurrence relations which generate these are often directly involved with continued fractions. See, for example, [2]. However, there was no good book on orthogonal polynomials when Ramanujan was working, and no one in England knew much about them when Ramanujan was there. Szegő started the serious development of orthogonal polynomials about the time Ramanujan died. It is a shame he was not aware of this subject, for it is a source of many beautiful identities Ramanujan would have loved, and would have given him another tool to find new results.

#### Acknowledgement

This work was supported in part by NSF grant DMS – 9300524.



## References

- [1] Berndt B C, Ramanujan's Notebooks, Part I, (1985) (New York: Springer)
- [2] Masson D, Wilson polynomials and some continued fractions of Ramanujan, *Rocky Mountain J. Math.*, **21** (1991) 489–499
- [3] Nikiforov A F, Suslov S K and Uvarov V B, Classical orthogonal polynomials of a discrete variable (1991) (Berlin: Springer)
- [4] Ramanujan S, Notebooks, volume 2, Tata Institute of Fundamental Research, Bombay, 1957
- [5] Srinivasa Rao K, Rajeswari V and King R C, Solutions of Diophantine equations and degree-one polynomial zeros of Racah coefficients, *J. Phys.* **A21** (1988) 1959–1070
- [6] Szegő G, *Orthogonal polynomials*, fourth edition, *Am. Math. Soc.*, Providence, RI, 1975
- [7] Tchebychef P L, Sur une nouvelle série, Oeuvres de P L, Tchebychef, I Chelsea, New York, 1961, 381–384



# Two remarkable doubly exponential series transformations of Ramanujan

BRUCE C BERNDT and JAMES LEE HAFNER\*

Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801, USA

\*IBM Research Division, Almaden Research Center, K53/802, 650 Harry Road, San Jose, CA 95120-6099, USA

Dedicated to the memory of Professor K G Ramanathan

**Abstract.** The purpose of this note is to prove two doubly exponential series transformations found in Ramanujan's second notebook.

**Keywords.** Doubly exponential series; Bernoulli numbers; gamma function; Poisson summation.

Ramanujan's notebooks [2] contain many fascinating theorems, which, it would seem, would never have been discovered by any other person. As excellent illustrations of this, Ramanujan offers transformation formulas for two doubly exponential series. It is very surprising that such elegant transformations exist. Although beautiful by themselves, we think that they will be useful in other investigations. These two remarkable series identities are stated without proof by Ramanujan on page 279 in his second notebook [2] and are numbered 4) and 5) on that page. The purpose of this note is to provide the missing proofs.

Ramanujan's two formulas can be stated as follows. First, if  $\alpha\beta = 2\pi$ , then

$$\alpha \sum_{k=0}^{\infty} \exp(-ne^{k\alpha}) = \alpha \left\{ \frac{1}{2} + \sum_{k=1}^{\infty} \frac{(-1)^{k-1} n^k}{k!(e^{k\alpha} - 1)} \right\} - \gamma - \log n + 2 \sum_{k=1}^{\infty} \varphi(k\beta), \quad (1)$$

where

$$\varphi(\beta) = \sqrt{\frac{\pi}{\beta \sinh(\pi\beta)}} \cos\left(\beta \log \frac{\beta}{n} - \beta - \frac{\pi}{4} - \frac{B_2}{1 \cdot 2\beta} + \dots\right). \quad (2)$$

Second, if  $\alpha\beta = \pi/2$  then

$$\begin{aligned} \alpha \sum_{k=0}^{\infty} (-1)^k \exp(-ne^{(2k+1)\alpha}) &= \alpha \left\{ \frac{1}{2} + \sum_{k=1}^{\infty} \frac{(-1)^k n^k}{k!(e^{k\alpha} + e^{-k\alpha})} \right\} \\ &\quad + \sum_{k=0}^{\infty} (-1)^k \psi((2k+1)\beta), \end{aligned} \quad (3)$$

where

$$\psi(\beta) = \sqrt{\frac{\pi}{\beta \sinh(\pi\beta)}} \sin\left(\beta \log \frac{\beta}{n} - \beta - \frac{\pi}{4} - \frac{B_2}{1 \cdot 2\beta} + \frac{B_4}{3 \cdot 4\beta^3} - \dots\right). \quad (4)$$

In (1),  $\gamma$  denotes Euler's constant, and in (2) and (4),  $B_k, k \geq 0$ , denotes the  $k$ th Bernoulli number defined by

$$\frac{z}{e^z - 1} = \sum_{k=0}^{\infty} \frac{B_k}{k!} z^k, \quad |z| < 2\pi.$$

We emphasize that we have altered Ramanujan's notation in (1)–(4). In particular, in Ramanujan's convention, all even indexed Bernoulli numbers are positive.

The definitions of  $\varphi(\beta)$  and  $\psi(\beta)$  given in (2) and (4), respectively, are decidedly enigmatic. Appearing in the arguments of the trigonometric functions are apparently asymptotic series as  $\beta$  tends to  $\infty$ . Thus the definitions of  $\varphi(\beta)$  and  $\psi(\beta)$  are imprecise, and so Ramanujan's claims are unclear. Nonetheless, we shall show that (1) and (3) are correct, if (2) and (4) are properly interpreted.

We begin by defining functions  $G(\beta)$  and  $B(\beta)$  by

$$\begin{aligned} \Gamma(i\beta + 1) &= (i\beta)^{i\beta + 1/2} \exp(-i\beta) \sqrt{2\pi} G(\beta) \\ &= (i\beta)^{i\beta + 1/2} \exp(-i\beta) \sqrt{2\pi} \exp\{-iB(\beta)\}. \end{aligned} \quad (5)$$

Then [4, pp. 252–253], as  $\beta$  tends to  $\infty$ ,

$$B(\beta) \sim \sum_{k=1}^{\infty} \frac{(-1)^{k-1} B_{2k}}{(2k-1)(2k)\beta^{2k-1}}$$

and

$$G(\beta) \sim 1 + \frac{1}{12i\beta} - \frac{1}{288\beta^2} - \frac{139}{51840(i\beta)^3} - \frac{571}{2488320\beta^4} + \dots \quad (6)$$

Ramanujan less explicitly gives the asymptotic expansion for  $B(\beta)$  in the argument to the trigonometric functions in (2) and (4).

We can now state our main results. Immediately after this, we will derive Ramanujan's identities as consequences.

**Theorem 1.** Let  $n$ ,  $\alpha$ , and  $\beta$  be positive with  $\alpha\beta = 2\pi$ . Then (1) holds, where

$$\begin{aligned} \varphi(\beta) &= \frac{1}{\beta} \operatorname{Im} \{ n^{-i\beta} \Gamma(i\beta + 1) \} \\ &= \sqrt{\frac{2\pi}{\beta}} \exp(-\pi\beta/2) \left\{ \sin\left(\beta \log \frac{\beta}{n} - \beta + \frac{\pi}{4}\right) \operatorname{Re}\{G(\beta)\} \right. \\ &\quad \left. + \cos\left(\beta \log \frac{\beta}{n} - \beta + \frac{\pi}{4}\right) \operatorname{Im}\{G(\beta)\} \right\} \\ &\sim \sqrt{\frac{2\pi}{\beta}} \exp(-\pi\beta/2) \left\{ \sin\left(\beta \log \frac{\beta}{n} - \beta + \frac{\pi}{4}\right) \left\{ 1 - \frac{1}{288\beta^2} + \dots \right\} \right. \\ &\quad \left. - \cos\left(\beta \log \frac{\beta}{n} - \beta + \frac{\pi}{4}\right) \left\{ \frac{1}{12\beta} + \dots \right\} \right\}, \end{aligned} \quad (7)$$

as  $\beta$  tends to  $\infty$ .

**Theorem 2.** Let  $n$ ,  $\alpha$ , and  $\beta$  be positive with  $\alpha\beta = \pi/2$ . Then (3) holds, where

$$\begin{aligned}\psi(\beta) &= \frac{1}{\beta} \operatorname{Re}\{n^{-i\beta} \Gamma(i\beta + 1)\} \\ &= -\sqrt{\frac{2\pi}{\beta}} \exp(-\pi\beta/2) \left\{ \cos\left(\beta \log \frac{\beta}{n} - \beta + \frac{\pi}{4}\right) \operatorname{Re}\{G(\beta)\} \right. \\ &\quad \left. - \sin\left(\beta \log \frac{\beta}{n} - \beta + \frac{\pi}{4}\right) \operatorname{Im}\{G(\beta)\} \right\} \\ &\sim -\sqrt{\frac{2\pi}{\beta}} \exp(-\pi\beta/2) \left\{ \cos\left(\beta \log \frac{\beta}{n} - \beta + \frac{\pi}{4}\right) \left\{ 1 - \frac{1}{288\beta^2} + \dots \right\} \right. \\ &\quad \left. + \sin\left(\beta \log \frac{\beta}{n} - \beta + \frac{\pi}{4}\right) \left\{ \frac{1}{12\beta} + \dots \right\} \right\},\end{aligned}\tag{8}$$

as  $\beta$  tends to  $\infty$ .

We first show that Ramanujan's definitions of (2) and (4) are compatible with the far right sides of (7) and (8), respectively. As  $\beta$  tends to  $\infty$ ,

$$\begin{aligned}&\sqrt{\frac{\pi}{\beta \sinh(\pi\beta)}} \cos\left(\beta \log \frac{\beta}{n} - \beta - \frac{\pi}{4} - B(\beta)\right) \\ &= \sqrt{\frac{2\pi}{\beta}} \exp(-\pi\beta/2) (1 - \exp(-2\pi\beta))^{-1/2} \sin\left(\beta \log \frac{\beta}{n} - \beta + \frac{\pi}{4} - B(\beta)\right) \\ &\sim \sqrt{\frac{2\pi}{\beta}} \exp(-\pi\beta/2) \left\{ \sin\left(\beta \log \frac{\beta}{n} - \beta + \frac{\pi}{4}\right) \cos B(\beta) \right. \\ &\quad \left. - \cos\left(\beta \log \frac{\beta}{n} - \beta + \frac{\pi}{4}\right) \sin B(\beta) \right\}.\end{aligned}$$

Thus (2) and (7) are in agreement. The argument showing that (4) and (8) agree is similar. This justifies Ramanujan's claims. We now proceed to prove the two main theorems.

*Proof (Theorem 1).* First,

$$\begin{aligned}\sum_{k=1}^{\infty} \frac{(-1)^{k-1} n^k}{k!(e^{k\alpha} - 1)} &= \sum_{k=1}^{\infty} \frac{(-1)^{k-1} n^k}{k! e^{k\alpha}} \sum_{j=0}^{\infty} e^{-kj\alpha} \\ &= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \frac{(-1)^{k-1} n^k e^{-kj\alpha}}{k!}\end{aligned}$$

Thus, the proposed identity may be written in the equivalent form

$$\begin{aligned} & \alpha \sum_{k=1}^{\infty} (e^{-ne^{k\alpha}} + e^{-ne^{-k\alpha}} - 1) - \frac{1}{2}\alpha + \alpha e^{-n} \\ &= -\gamma - \log n + 2 \sum_{k=1}^{\infty} \varphi(k\beta). \end{aligned} \quad (9)$$

Secondly, we apply Poisson's summation formula [3, p. 60] to the function

$$f(x) := \exp(-ne^x) + \exp(-ne^{-x}) - 1. \quad (10)$$

Observing that  $f(0) = 2\exp(-n) - 1$ , we find that, for  $\alpha, \beta > 0$  with  $\alpha\beta = 2\pi$ ,

$$\begin{aligned} & \alpha \left\{ \frac{1}{2}(2e^{-n} - 1) + \sum_{k=1}^{\infty} (\exp(-ne^{k\alpha}) + \exp(-ne^{-k\alpha}) - 1) \right\} \\ &= \int_0^{\infty} f(x) dx + 2 \sum_{k=1}^{\infty} \int_0^{\infty} f(x) \cos(k\beta x) dx. \end{aligned} \quad (11)$$

Comparing (9) and (11), we see that it remains to prove that

$$-\gamma - \log n + 2 \sum_{k=1}^{\infty} \varphi(k\beta) = \int_0^{\infty} f(x) dx + 2 \sum_{k=1}^{\infty} \int_0^{\infty} f(x) \cos(k\beta x) dx. \quad (12)$$

Observe by (10) that  $f(x)$  is even. Setting  $u = e^x$ , we find that

$$\begin{aligned} \int_0^{\infty} f(x) dx &= \frac{1}{2} \int_{-\infty}^{\infty} f(x) dx \\ &= \frac{1}{2} \int_0^{\infty} (e^{-nu} + e^{-n/u} - 1) \frac{du}{u} \\ &= \frac{1}{2} \left( - \int_0^{1/n} \frac{1 - e^{-nu}}{u} du + \int_{1/n}^{\infty} \frac{e^{-nu}}{u} du \right. \\ &\quad \left. + \int_0^{1/n} \frac{e^{-n/u}}{u} du - \int_{1/n}^{\infty} \frac{1 - e^{-n/u}}{u} du \right) \\ &= \frac{1}{2} \left( - \int_0^1 \frac{1 - e^{-x}}{x} dx + \int_1^{\infty} \frac{e^{-x}}{x} dx \right. \\ &\quad \left. + \int_n^{\infty} \frac{e^{-x}}{x} dx - \int_0^{n^2} \frac{1 - e^{-x}}{x} dx \right). \end{aligned}$$

Since [1, p. 103]

$$\gamma = \int_0^1 \frac{1 - e^{-x}}{x} dx - \int_1^{\infty} \frac{e^{-x}}{x} dx,$$

we find that

$$\begin{aligned}
 \int_0^{\infty} f(x) dx &= \frac{1}{2} \left( -\gamma + \int_1^{\infty} \frac{e^{-x}}{x} dx - \int_0^1 \frac{1 - e^{-x}}{x} dx - \int_0^{n^2} \frac{dx}{x} \right) \\
 &= \frac{1}{2} (-\gamma - \gamma - \log n^2) \\
 &= -\gamma - \log n.
 \end{aligned} \tag{13}$$

Using (13) in (12), we find that it suffices to prove that

$$\sum_{k=1}^{\infty} \varphi(k\beta) = \sum_{k=1}^{\infty} \int_0^{\infty} f(x) \cos(k\beta x) dx. \tag{14}$$

Set

$$I := I(\beta) := \int_0^{\infty} f(x) \cos(\beta x) dx.$$

By (14), it now suffices to prove that  $I(\beta) = \varphi(\beta)$ , where  $\varphi(\beta)$  is defined by (7). Letting  $u = e^x$ , we find that

$$\begin{aligned}
 I &= \frac{1}{2} \int_{-\infty}^{\infty} f(x) \cos(\beta x) dx \\
 &= \frac{1}{2} \int_0^{\infty} (e^{-nu} + e^{-n/u} - 1) \cos(\beta \log u) \frac{du}{u}.
 \end{aligned}$$

Integrating by parts, we find that

$$\begin{aligned}
 I &= \frac{n}{2\beta} \int_0^{\infty} \left( e^{-nu} - \frac{1}{u^2} e^{-n/u} \right) \sin(\beta \log u) du \\
 &= \frac{n}{2\beta} \left( \int_0^{\infty} e^{-nu} \sin(\beta \log u) du - \int_0^{\infty} \frac{e^{-n/u}}{u^2} \sin(\beta \log u) du \right) \\
 &= \frac{n}{2\beta} (I_1 - I_2),
 \end{aligned}$$

say. Setting  $t = 1/u$  in  $I_2$ , we deduce that  $I_2 = -I_1$ . Hence,

$$\begin{aligned}
 I &= \frac{n}{\beta} I_1 \\
 &= \frac{n}{\beta} \int_0^{\infty} e^{-nu} \sin(\beta \log u) du \\
 &= \frac{n}{2\beta i} \int_0^{\infty} (e^{-nu} u^{i\beta} - e^{-nu} u^{-i\beta}) du \\
 &= \frac{1}{2\beta i} \{ n^{-i\beta} \Gamma(i\beta + 1) - n^{i\beta} \Gamma(-i\beta + 1) \}
 \end{aligned}$$

$$= \frac{1}{\beta} \operatorname{Im} \{ n^{-i\beta} \Gamma(i\beta + 1) \}.$$

Hence,  $I = I(\beta) = \varphi(\beta)$ , by (7). This completes the proof of (1).

Lastly, from (5),

$$\begin{aligned} \varphi(\beta) &= \frac{1}{\beta} \operatorname{Im} \left( (i\beta)^{1/2} \left( \frac{i\beta}{ne} \right)^{i\beta} \sqrt{2\pi} G(\beta) \right) \\ &= \sqrt{\frac{2\pi}{\beta}} \operatorname{Im} (e^{i(\pi/4 + \beta \log(i\beta/(ne)))} G(\beta)) \\ &= \sqrt{\frac{2\pi}{\beta}} e^{-\pi\beta/2} \operatorname{Im} (e^{i(\pi/4 + \beta \log(\beta/n) - \beta)} G(\beta)). \end{aligned}$$

Hence, the second equality in (7) follows, and the asymptotic formula follows by using (6).  $\square$

*Proof (Theorem 2).* First,

$$\begin{aligned} \sum_{k=1}^{\infty} \frac{(-1)^k n^k}{k!(e^{k\alpha} + e^{-k\alpha})} &= \sum_{k=1}^{\infty} \frac{(-1)^k n^k}{k! e^{k\alpha}} \sum_{j=0}^{\infty} (-1)^j e^{-2kj\alpha} \\ &= \sum_{j=0}^{\infty} (-1)^j (\exp(-ne^{-(2j+1)\alpha}) - 1). \end{aligned}$$

Thus, the proposed identity (3) can be recast in the form

$$\begin{aligned} \alpha \sum_{k=0}^{\infty} (-1)^k (\exp(-ne^{(2k+1)\alpha}) - \exp(-ne^{-(2k+1)\alpha}) + 1) \\ = \frac{1}{2}\alpha + \sum_{k=0}^{\infty} (-1)^k \psi((2k+1)\beta). \end{aligned} \quad (15)$$

Next, we apply the Poisson summation formula for Fourier sine transforms [3, p. 66] to the function

$$f(x) := \exp(-ne^x) - \exp(-ne^{-x}) + 1.$$

Thus, for  $\alpha, \beta > 0$  and  $\alpha\beta = \pi/2$ ,

$$\begin{aligned} \alpha \sum_{k=0}^{\infty} (-1)^k (\exp(-ne^{(2k+1)\alpha}) - \exp(-ne^{-(2k+1)\alpha}) + 1) \\ = \sum_{k=0}^{\infty} (-1)^k \int_0^{\infty} f(x) \sin((2k+1)\beta x) dx. \end{aligned} \quad (16)$$

We recall that

$$\frac{\alpha}{2} = \frac{\pi}{4\beta} = \frac{1}{\beta} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1}. \quad (17)$$



$$\begin{aligned}
& \sum_{k=0}^{\infty} (-1)^k \psi((2k+1)\beta) \\
&= \sum_{k=0}^{\infty} (-1)^k \left( \int_0^{\infty} f(x) \sin((2k+1)\beta x) dx - \frac{1}{(2k+1)\beta} \right). \tag{18}
\end{aligned}$$

Set

$$I := I(\beta) := \int_0^{\infty} f(x) \sin(\beta x) dx - \frac{1}{\beta}.$$

By (18), we now see that it suffices to prove that  $I(\beta) = \psi(\beta)$ , where  $\psi(\beta)$  is defined by (8).

Setting  $x = e^u$  and integrating by parts, we find that

$$\begin{aligned}
I &= \int_0^{\infty} (\exp(-ne^x) - \exp(-ne^{-x}) + 1) \sin(\beta x) dx - \frac{1}{\beta} \\
&= \int_0^{\infty} (e^{-nu} - e^{-n/u} + 1) \sin(\beta \log u) \frac{du}{u} - \frac{1}{\beta} \\
&= -\frac{n}{\beta} \int_1^{\infty} \left( e^{-nu} + \frac{1}{u^2} e^{-n/u} \right) \cos(\beta \log u) du \\
&= -\frac{n}{\beta} \int_0^1 \left( e^{-nt} + \frac{1}{t^2} e^{-n/t} \right) \cos(\beta \log t) dt,
\end{aligned}$$

where we set  $u = 1/t$ . Hence,

$$\begin{aligned}
I &= -\frac{n}{2\beta} \int_0^{\infty} \left( e^{-nt} + \frac{1}{t^2} e^{-n/t} \right) \cos(\beta \log t) dt \\
&= -\frac{n}{2\beta} (I_1 + I_2),
\end{aligned}$$

say. Letting  $t = 1/u$  in  $I_2$ , we easily find that  $I_2 = I_1$ . Consequently,

$$\begin{aligned}
I &= -\frac{n}{\beta} I_1 \\
&= -\frac{n}{\beta} \int_0^{\infty} e^{-nt} \cos(\beta \log t) dt \\
&= -\frac{n}{2\beta} \int_0^{\infty} (e^{-nt} t^{i\beta} + e^{-nt} t^{-i\beta}) dt \\
&= -\frac{1}{2\beta} \{ n^{-i\beta} \Gamma(i\beta + 1) + n^{i\beta} \Gamma(-i\beta + 1) \} \\
&= -\frac{1}{\beta} \operatorname{Re} \{ n^{-i\beta} \Gamma(i\beta + 1) \}.
\end{aligned}$$

Thus, we have shown that  $I(\beta) = \psi(\beta)$ , by (8). This completes the proof of (3).

The remaining two claims in (8) follow as in the proof of Theorem 1. □

## References

- [1] Berndt B C, *Ramanujan's Notebooks, Part I*, (New York: Springer-Verlag) (1985)
- [2] Ramanujan S, *Notebooks* (2 volumes) (Bombay: Tata Institute of Fundamental Research) (1957)
- [3] Titchmarsh E C, *Introduction to the theory of Fourier integrals*, (2nd ed., Oxford: Clarendon Press) (1948)
- [4] Whittaker E T and Watson G N, *A course of modern analysis*, 4th ed., (Cambridge: University Press) (1966)

## Kolmogorov's existence theorem for Markov processes in $C^*$ algebras

B V RAJARAMA BHAT and K R PARTHASARATHY

Indian Statistical Institute, 7, S.J.S. Sansanwal Marg, New Delhi 110016, India

Dedicated to the memory of Professor K G Ramanathan

**Abstract.** Given a family of transition probability functions between measure spaces and an initial distribution Kolmogorov's existence theorem associates a unique Markov process on the product space. Here a canonical non-commutative analogue of this result is established for families of completely positive maps between  $C^*$  algebras satisfying the Chapman-Kolmogorov equations. This could be the starting point for a theory of quantum Markov processes.

**Keywords.** Completely positive map; Markov process; GNS principle.

### 1. Introduction

Let  $(X_i, \mathcal{F}_i)$ ,  $i = 0, 1, 2, \dots$  be Polish measurable spaces and let  $P_i(x_i, dx_{i+1})$  be a transition probability from  $(X_i, \mathcal{F}_i)$  to  $(X_{i+1}, \mathcal{F}_{i+1})$  for each  $i$ . Given a probability measure  $\mu$  on  $(X_0, \mathcal{F}_0)$  it follows from Kolmogorov's extension theorem that there exists a unique probability measure  $P_\mu$  on the infinite product space  $(\Omega, \mathcal{F}) = \bigotimes_{i=0}^{\infty} (X_i, \mathcal{F}_i)$

such that, for every finite  $n$ , its projection or marginal distribution  $P_\mu^n$  in  $\bigotimes_{i=0}^n (X_i, \mathcal{F}_i)$  is given by

$$P_\mu^n(E_0 \times E_1 \times \dots \times E_n) = \int_{E_0 \times E_1 \times \dots \times E_n} \mu(dx_0) P_0(x_0, dx_1) P_1(x_1, dx_2) \dots P_n(x_{n-1}, dx_n) \quad (1.1)$$

for all  $E_i \in \mathcal{F}_i$ ,  $i = 0, 1, 2, \dots, n$ . The probability space  $(\Omega, \mathcal{F}, P_\mu)$  describes the Markov process with initial distribution  $\mu$  and transition probability  $P_i(\cdot, \cdot)$  for transition from a state at time  $i$  to a new state at time  $i + 1$ . This can be described in a  $*$  algebraic language as follows. Denote by  $\mathcal{A}_i$  the commutative  $*$  algebra of all complex valued bounded measurable functions on  $(X_i, \mathcal{F}_i)$ . Introduce the positive unital operator  $T(i, i + 1): \mathcal{A}_{i+1} \rightarrow \mathcal{A}_i$  by

$$(T(i, i + 1)g)(x_i) = \int g(x_{i+1}) P_i(x_i, dx_{i+1}).$$

For any  $i \leq k$  define  $T(i, k): \mathcal{A}_k \rightarrow \mathcal{A}_i$  by

$$T(i, k) = \begin{cases} \text{identity} & \text{if } i = k, \\ T(i, i + 1) T(i + 1, i + 2) \dots T(k - 1, k) & \text{if } i < k. \end{cases}$$

The family  $\{T(i, k), i \leq k\}$  of transition operators obeys the Chapman-Kolmogorov equations:

$$T(i, k) T(k, \ell) = T(i, \ell) \quad \text{for } i \leq k \leq \ell.$$

Let  $\mathcal{H}$  be the Hilbert space  $L^2(P_\mu)$  and  $F(i)$  denote the Hilbert space projection on the subspace of functions depending only on the first  $i + 1$  coordinates  $(x_0, x_1, \dots, x_i)$  of  $\omega = (x_0, x_1, x_2, \dots)$  in  $\Omega$ . Then  $\{F(i)\}$  is an increasing sequence of projections in  $\mathcal{H}$ . For any  $g \in \mathcal{A}_i$  define the operator  $j_i(g)$  in  $\mathcal{H}$  by

$$(j_i(g)\phi)(\omega) = g(x_i)(F(i)\phi)(\omega), \quad \omega = (x_0, x_1, \dots).$$

Then  $j_i$  is a  $*$  homomorphism from  $\mathcal{A}_i$  into the  $*$  algebra  $\mathcal{B}(\mathcal{H})$  of all bounded operators in  $\mathcal{H}$ . The Markov property of the stochastic process  $(\Omega, \mathcal{F}, P_\mu)$  is encapsulated in the operator relations

$$j_k(1) = F(k), \tag{1.2}$$

$$F(i)j_k(g)F(i) = j_i(T(i, k)g), \quad g \in \mathcal{A}_k, \quad i \leq k. \tag{1.3}$$

The relations (1.1) can be expressed as

$$\begin{aligned} & \langle u, j_0(g_0)j_1(g_1) \cdots j_n(g_n)v \rangle \\ &= \int (\bar{u}vg_0)(x_0)g_1(x_1) \cdots g_n(x_n)dP_\mu(\omega) \end{aligned} \tag{1.4}$$

for all  $u, v$  in the range of  $F(0)$  and  $g_i \in \mathcal{A}_i$ ,  $i = 0, 1, 2, \dots, n$ . Here  $\omega$  denotes the sequence  $(x_0, x_1, \dots)$ . We may call the triple  $(\mathcal{H}, F, j_k, k = 0, 1, 2, \dots)$  consisting of the Hilbert space  $\mathcal{H}$ , the filtration of projections  $F(k)$  increasing in  $k$  and the family  $\{j_k, k = 0, 1, 2, \dots\}$  of  $*$  (but nonunital) homomorphisms, a Markov process with transition operators  $\{T(i, j), i \leq j\}$ . A similar description of a Markov process in continuous time is also possible.

In the context of quantum or non-commutative probability theory there have been several partial attempts (for example, by Accardi, Frigerio and Lewis [AFL], Emch [E], Sauvageot [S] and Vincent-Smith [Vi-S]) to construct Markov processes when transition probabilities between measurable spaces, or equivalently, the transition operators between the corresponding commutative  $*$  algebras of bounded measurable functions are replaced by unital and completely positive linear maps between unital  $*$  algebras of operators in Hilbert spaces. In the present paper we shall start with a family of completely positive maps between  $C^*$  algebras which obey the Chapman-Kolmogorov equations and build a unique canonical minimal Markov process, using the GNS principle. Rather remarkably, this minimal process, when restricted to the centres of the different  $C^*$  algebras that are involved, can be obtained as a conditional expectation of a completely commutative process. The definition of a Markov process that we shall adopt is inspired by the equations (1.2)–(1.4).

## 2. The basic construction

Let  $\mathcal{A}_t$  be a unital  $C^*$  algebra of bounded operators in a complex Hilbert space  $\mathcal{H}_t$ , for every  $t \geq 0$ . The time index  $t$  here may be discrete or continuous. It is useful to

imagine any hermitian element  $x \in \mathcal{A}_t$  as a real valued observable concerning a system at time  $t$ . For every  $0 \leq s \leq t < \infty$  let  $T(s, t): \mathcal{A}_t \rightarrow \mathcal{A}_s$  be a linear, unital and completely positive map (hereafter called simply a c.p. map) satisfying the following: (i)  $T(s, s)$  is the identity map on  $\mathcal{A}_s$ ; (ii)  $T(r, t) = T(r, s) T(s, t)$  for all  $0 \leq r \leq s \leq t < \infty$ . When (i) and (ii) hold we say that the family  $\{T(s, t)\}$  of c.p. maps obeys the Chapman-Kolmogorov equations and call it a family of *transition operators*. Complete positivity is equivalent to the condition

$$\sum_{i,j} X_i^* \{T(s, t)(Y_i^* Y_j)\} X_j \geq 0$$

for all bounded operators  $X_i$  in  $\mathcal{K}_s$  and elements  $Y_i \in \mathcal{A}_t$ , the summation being over any finite index set. Another equivalent description of complete positivity is that, for every finite  $n$ , the matrix  $((T(s, t)(Y_{ij})))_{1 \leq i, j \leq n}$ , viewed as an operator in the  $n$ -fold direct sum  $\mathcal{K}_s \oplus \dots \oplus \mathcal{K}_s$ , is positive whenever  $((Y_{ij}))_{1 \leq i, j \leq n}$  is positive in the  $n$ -fold direct sum  $\mathcal{K}_t \oplus \dots \oplus \mathcal{K}_t$  with  $Y_{ij} \in \mathcal{A}_t$  for each  $i, j$ .

Denote by  $\Gamma_0(\mathbb{R}_+) = \Gamma_0$  the set  $\{\sigma \mid \sigma \subset \mathbb{R}_+, 0 \in \sigma, \#\sigma < \infty\}$ , where  $\#\sigma$  denotes the cardinality of  $\sigma$ . When  $\#\sigma = n$  and  $t_i \in \sigma$ ,  $i = 1, 2, \dots, n$  are distinct we always express it as  $\sigma = \{t_1, t_2, \dots, t_n\}$  with  $t_1 > t_2 > \dots > t_n = 0$ . When  $X_{t_i} \in \mathcal{A}_{t_i}$  for each  $i = 1, 2, \dots, n$  we denote the  $n$ -length sequence  $\{X_{t_1}, X_{t_2}, \dots, X_{t_n}\}$  by  $X(\sigma)$ . Suppose that  $\sigma = \{s_1, s_2, \dots, s_m\}$ ,  $\delta = \{t_1, t_2, \dots, t_n\}$  and  $\sigma \cup \delta = \{r_1, r_2, \dots, r_k\}$  are in  $\Gamma_0$ . For any  $X(\sigma)$  with  $X_{s_i} \in \mathcal{A}_{s_i}$  we write  $X^\sigma(\sigma \cup \delta)$  for the sequence  $Y(\sigma \cup \delta)$  defined by

$$Y_{r_i} = \begin{cases} X_{s_j} & \text{if } r_i = s_j \text{ for some } j = 1, 2, \dots, n, \\ I_{r_i} & \text{otherwise,} \end{cases}$$

where  $I_r$  is the identity element in  $\mathcal{A}_r$ . Denote by  $\tilde{A}$  the set of all sequences of the form  $X(\sigma)$  with  $\sigma$  varying in  $\Gamma_0$  and write

$$\mathcal{M} = \tilde{A} \times \mathcal{K}_0, \quad (2.1)$$

$$\mathcal{M}_t = \begin{cases} \{(X(\sigma), u) \in \mathcal{M}, \sigma = (t, t_2, \dots, t_n), n = 2, 3, \dots\} & \text{if } t > 0 \\ \mathcal{A}_0 \times \mathcal{K}_0 & \text{if } t = 0 \end{cases} \quad (2.2)$$

To the family  $\{T(s, t)\}$  of transition operators we now associate a function  $L_T$  on the set  $\mathcal{M} \times \mathcal{M}$  as follows:

$$\begin{aligned} L_T((X(\sigma), u), (Y(\delta), v)) &= \langle u, X_0^* \{T(0, t_{n-1})(X_{t_{n-1}}^* \{T(t_{n-1}, t_{n-2}) \\ &\quad (\dots X_{t_2}^* \{T(t_2, t_1)(X_{t_1}^* Y_{t_1})\} Y_{t_2} \dots)\} Y_{t_{n-1}}) Y_0\} v \rangle \\ &\quad \text{if } \sigma = \{t_1, t_2, \dots, t_n\}, \end{aligned} \quad (2.3)$$

and

$$L_T((X(\sigma), u), (Y(\delta), v)) = L_T((X^\sigma(\sigma \cup \delta), u), (Y^\delta(\sigma \cup \delta), v)). \quad (2.4)$$

### PROPOSITION 2.1.

$L_T$  is a positive definite kernel on  $\mathcal{M} \times \mathcal{M}$ , i.e., for any  $n = 1, 2, \dots$ , complex scalars  $c_i$  and elements  $(X_i(\sigma_i), u_i) \in \mathcal{M}$ ,  $i = 1, 2, \dots, n$  the following inequality holds:

$$\sum_{1 \leq i, j \leq n} \bar{c}_i c_j L_T((X_i(\sigma_i), u_i), (X_j(\sigma_j), u_j)) \geq 0 \quad (2.5)$$

$$L_T((X(\sigma), u), (Y(\sigma), v)) = L_T((X^\sigma(\sigma \cup \delta), u), (Y^\delta(\sigma \cup \delta), v)). \quad (2.6)$$

It suffices to prove this relation when  $\delta = \{t, 0\}$ ,  $\sigma = \{t_1, t_2, \dots, t_{n-1}, 0\}$ ,  $t \neq t_i$  for every  $i$ , since the more general case would follow by induction. In this special case (2.6) follows easily from (2.3) with  $\sigma$  replaced by  $\sigma \cup \delta$  and the Chapman-Kolmogorov equations. In view of (2.4) it is enough to prove (2.5) when  $\sigma_i = \sigma$  for each  $i$ , for otherwise, we may replace all the  $\sigma_i$ 's by  $\sigma = \bigcup_i \sigma_i$ . Let  $\sigma = \{t_1, t_2, \dots, t_{m-1}, t_m = 0\}$  and

$$X_i(\sigma) = (X_{it_1}, X_{it_2}, \dots, X_{it_m}), \quad i = 1, 2, \dots, n.$$

Define inductively the following operators:

$$\begin{aligned} Z_{ij}(t_1) &= X_{it_1}^* X_{jt_1} \\ Z_{ij}(t_r) &= X_{it_r}^* T(t_r, t_{r-1})(Z_{ij}(t_{r-1})) X_{jt_r}, \\ r &= 2, 3, \dots, m. \end{aligned}$$

Clearly, the matrix  $((Z_{ij}(t_1)))$  is a positive operator in the  $n$ -fold direct sum  $\mathcal{H}_{t_1} \oplus \dots \oplus \mathcal{H}_{t_1}$ . If  $((Z_{ij}(t_{r-1})))$  is a positive operator in  $\mathcal{H}_{t_{r-1}} \oplus \dots \oplus \mathcal{H}_{t_{r-1}}$ , the complete positivity of  $T(t_r, t_{r-1})$  implies that  $((Z_{ij}(t_r)))$  is positive in  $\mathcal{H}_{t_r} \oplus \dots \oplus \mathcal{H}_{t_r}$ . Thus, by induction,  $((Z_{ij}(t_m)))$  is a positive operator in  $\mathcal{H}_0 \oplus \dots \oplus \mathcal{H}_0$ . If we write  $\xi = \bigoplus_{i=1}^n c_i u_i$  in  $\mathcal{H}_0 \oplus \dots \oplus \mathcal{H}_0$  we have

$$\sum_{1 \leq i, j \leq n} \bar{c}_i c_j L_T((X_i(\sigma), u_i), (X_j(\sigma), u_j)) = \langle \xi, ((Z_{ij}(t_m))) \xi \rangle \geq 0. \quad \blacksquare$$

## PROPOSITION 2.2.

There exists a Hilbert space  $\mathcal{H}$  and a map  $\lambda: \mathcal{M} \rightarrow \mathcal{H}$  satisfying the following:

- (i)  $\langle \lambda(X(\sigma), u), \lambda(Y(\delta), v) \rangle \equiv L_T((X(\sigma), u), (Y(\delta), v))$ ;
- (ii) The set  $\{\lambda(X(\sigma), u) | (X(\sigma), u) \in \mathcal{M}\}$  is total in  $\mathcal{H}$ ;
- (iii) If  $\mathcal{H}'$  is another Hilbert space and  $\lambda': \mathcal{M} \rightarrow \mathcal{H}'$  satisfying (i) and (ii) with  $(\mathcal{H}, \lambda)$  replaced by  $(\mathcal{H}', \lambda')$  then there exists a unitary operator  $W: \mathcal{H} \rightarrow \mathcal{H}'$  such that  $W \circ \lambda = \lambda'$ ;
- (iv)  $\lambda((X(\sigma), u)) = \lambda(X^\sigma(\sigma \cup \delta), u)$  for all  $(X(\sigma), u) \in \mathcal{M}$  and  $\delta \in \Gamma_0$ .

*Proof.* (i), (ii) and (iii) are immediate from Proposition 2.1 and the G.N.S. principle. (See, for example, Proposition 15.4, [P]). By (2.3) and (2.4) we have

$$\begin{aligned} L_T((X(\sigma), u), (X(\sigma), u)) &= L_T((X(\sigma), u), (X^\sigma(\sigma \cup \delta), u)) \\ &= L_T((X^\sigma(\sigma \cup \delta), u), (X^\sigma(\sigma \cup \delta), u)) \end{aligned}$$

and hence by (i) in the proposition

$$\begin{aligned} \|\lambda(X(\sigma), u) - \lambda(X^\sigma(\sigma \cup \delta), u)\|^2 &= \|\lambda(X(\sigma), u)\|^2 + \|\lambda(X^\sigma(\sigma \cup \delta), u)\|^2 \\ &\quad - 2 \operatorname{Re} \langle \lambda(X(\sigma), u), \lambda(X^\sigma(\sigma \cup \delta), u) \rangle = 0. \quad \blacksquare \end{aligned}$$

*Remark.* When  $\sigma = \{t_1, t_2, \dots, t_n\}$  is fixed it is a consequence of (i) in Proposition 2.2 that  $\lambda((X_{t_1}, X_{t_2}, \dots, X_{t_n}), u)$  is multilinear on  $\mathcal{A}_{t_1} \times \dots \times \mathcal{A}_{t_n} \times \mathcal{H}_0$ .

### PROPOSITION 2.3.

In Proposition 2.2 let  $\mathcal{H}_t$  be the closed linear span of the set  $\{\lambda(X(\sigma), u) | (X(\sigma), u) \in \mathcal{M}_t\}$  where  $\mathcal{M}_t$  is defined by (2.1) and (2.2). Then  $\{\mathcal{H}_t, t \geq 0\}$  is an increasing family of subspaces of  $\mathcal{H}$  and the map  $V: u \rightarrow \lambda(I_0, u)$  is a unitary operator from  $\mathcal{H}_0$  to  $\mathcal{H}_0$ .

*Proof.* Let  $0 \leq s < t < \infty$ . Suppose  $\sigma = \{s, s_2, \dots, s_m\}$ . Then by property (iv) in Proposition 2.2 we have

$$\lambda((X_s, X_{s_2}, \dots, X_{s_m}), u) = \lambda((I_t, X_s, X_{s_2}, \dots, X_{s_m}), u)$$

and the right hand side belongs to  $\mathcal{H}_t$  by definition. This proves the first part. To prove the second part we first observe that

$$\langle \lambda(I_0, u), \lambda(I_0, v) \rangle_{\mathcal{H}} = \langle u, v \rangle_{\mathcal{H}_0}.$$

Thus  $V$  is an isometry from  $\mathcal{H}_0$  into  $\mathcal{H}_0$ . Furthermore (2.3) implies

$$\begin{aligned} & \|\lambda(X_0, u) - \lambda(I_0, X_0 u)\|^2 \\ &= L_T((X_0, u), (X_0, u)) + L_T((I_0, X_0 u), (I_0, X_0 u)) \\ &\quad - 2\operatorname{Re} L_T((X_0, u), (I_0, X_0 u)) \\ &= \langle u, X_0^* X_0 u \rangle + \langle X_0 u, X_0 u \rangle \\ &\quad - 2\operatorname{Re} \langle u, X_0^* (X_0 u) \rangle = 0. \end{aligned}$$

For any Hilbert space  $\mathcal{H}$  we denote by  $\mathcal{B}(\mathcal{H})$  the  $C^*$  algebra of all bounded operators on  $\mathcal{H}$ .

### PROPOSITION 2.4.

Let  $\mathcal{H}$ ,  $\mathcal{H}_t$ ,  $\lambda$ ,  $V$  be as in Proposition 2.3. Then there exists a unique  $*$  unital homomorphism  $j_t^0: \mathcal{A}_t \rightarrow \mathcal{B}(\mathcal{H}_t)$  for every  $t \geq 0$  satisfying the relations:

$$j_t^0(Y)\lambda((X_t, X_{t_2}, \dots, X_{t_n}), u) = \lambda((YX_t, X_{t_2}, \dots, X_{t_n}), u) \quad (2.7)$$

for all  $Y \in \mathcal{A}_t$ ,  $t > t_2 > \dots > t_n = 0$ ,  $u \in \mathcal{H}_0$ . Furthermore

$$V^* j_0^0(X) V = X \quad \text{for all } X \in \mathcal{A}_0.$$

*Proof.* Let  $Y \in \mathcal{A}_t$  be unitary. By (2.3) and the fact that  $\{T(s, t)\}$  is a family of transition operators it follows immediately that

$$\begin{aligned} & \langle \lambda((YX_t, X_{t_2}, \dots, X_{t_n}), u), \lambda((YZ_t, Z_{t_2}, \dots, Z_{t_n}), v) \rangle \\ &= L_T(((YX_t, X_{t_2}, \dots, X_{t_n}), u), ((YZ_t, Z_{t_2}, \dots, Z_{t_n}), v)) \\ &= L_T(((X_t, X_{t_2}, \dots, X_{t_n}), u), ((Z_t, Z_{t_2}, \dots, Z_{t_n}), v)) \\ &= \langle \lambda((X_t, X_{t_2}, \dots, X_{t_n}), u), \lambda((Z_t, Z_{t_2}, \dots, Z_{t_n}), v) \rangle \end{aligned}$$

for all  $X_t, Y_t \in \mathcal{A}_t, X_{t_1}, Y_{t_1} \in \mathcal{A}_{t_1}, u, v \in \mathcal{H}_0$ . This together with property (iv) of Proposition 2.2 implies that

$$\begin{aligned} & \langle \lambda(YX_t, X_{t_2}, \dots, X_{t_n}), u \rangle, \lambda(YZ_t, Z_{t_1}, Z_{t_2}, \dots, Z_{t_n}), v \rangle \\ &= \langle \lambda(X_t, X_{t_2}, \dots, X_{t_n}), u \rangle, \lambda(Z_t, Z_{t_1}, Z_{t_2}, \dots, Z_{t_n}), v \rangle \end{aligned}$$

Thus for any unitary  $Y$  in  $\mathcal{A}_t$  there exists a unitary operator  $j_t^0(Y)$  in  $\mathcal{H}_t$  satisfying (2.7). If  $Y_1, Y_2$  are unitary elements in  $\mathcal{A}_t$  it follows from the definitions that  $j_t^0(Y_1)j_t^0(Y_2) = j_t^0(Y_1 Y_2)$ . Since  $\lambda((X_t, X_{t_1}, \dots, X_{t_n}), u)$  is linear in the variable  $X_t$  and any element in  $\mathcal{A}_t$  is a linear combination of at most four unitary elements in  $\mathcal{A}_t$ , it follows that  $j_t^0(\cdot)$  defined for unitary elements extends linearly to  $\mathcal{A}_t$  as a \* unital homomorphism from  $\mathcal{A}_t$  into  $\mathcal{B}(\mathcal{H}_t)$ . The uniqueness part is obvious. To prove the last part we have to only note that by the definition of  $V$  in Proposition 2.3 and the last part of its proof

$$\begin{aligned} j_0^0(X)Vu &= j_0^0(X)\lambda(I_0, u) = \lambda(X, u) \\ &= \lambda(I_0, Xu) = VXu \end{aligned}$$

for all  $u \in \mathcal{H}_0$ . ■

**Theorem 2.5.** Let  $\mathcal{A}_t$  be a unital  $C^*$  algebra of operators in a Hilbert space  $\mathcal{H}_t$  for every  $t \geq 0$  and let  $T(s, t): \mathcal{A}_t \rightarrow \mathcal{A}_s, s \leq t$  be a family of transition operators. Then there exists a Hilbert space  $\mathcal{H}$ , an increasing family  $\{F(t), t \geq 0\}$  of projection operators on  $\mathcal{H}$ , a family of contractive \* homomorphisms  $j_t: \mathcal{A}_t \rightarrow \mathcal{B}(\mathcal{H}), t \geq 0$  and a unitary isomorphism  $V$  from  $\mathcal{H}_0$  onto the range of  $F(0)$  satisfying the following:

- (i)  $j_t(I_t) = F(t)$ ,  $I_t$  being the identity operator in  $\mathcal{H}_t$ ;
- (ii) for any  $0 \leq s \leq t < \infty, X \in \mathcal{A}_t$

$$F(s)j_t(X)F(s) = j_s(T(s, t)(X));$$

- (iii) the set  $\{j_{t_1}(X_1) \cdots j_{t_n}(X_n)Vu, t_1 > t_2 > \cdots > t_n = 0, X_i \in \mathcal{A}_{t_i} \text{ for each } i, n = 1, 2, \dots, u \in \mathcal{H}_0\}$  is total in  $\mathcal{H}$ ;
- (iv)  $j_0(X)V = VX$  for all  $X \in \mathcal{A}_0$  and for any  $u, v \in \mathcal{H}_0, \sigma = \{s_1 > s_2 > \cdots > s_m = 0\}, \delta = \{t_1 > t_2 > \cdots > t_n = 0\}$ ,

$$\begin{aligned} & X_i \in \mathcal{A}_{s_i}, Y_j \in \mathcal{A}_{t_j}, i = 1, 2, \dots, m, j = 1, 2, \dots, n \\ & \langle j_{s_1}(X_1)j_{s_2}(X_2) \cdots j_{s_m}(X_m)Vu, j_{t_1}(Y_1)j_{t_2}(Y_2) \cdots j_{t_n}(Y_n)Vv \rangle \\ &= L_T((X(\sigma), u), (Y(\delta), v)), \end{aligned}$$

where  $L_T$  is given by (2.3) and (2.4).

*Proof.* Let  $\mathcal{H}, \mathcal{H}_t, \lambda, V$  and  $j_t^0$  be as in Proposition 2.4. Define  $F(t)$  to be the projection on the subspace  $\mathcal{H}_t$ . By Proposition 2.3,  $F(t)$  is increasing in  $t$ . Define, for any  $X \in \mathcal{A}_t$ , the operator  $j_t(X)$  in  $\mathcal{H}$  by

$$j_t(X) = j_t^0(X)F(t) \quad \text{for any } t \geq 0.$$

Since  $j_t^0$  is a \* unital homomorphism from  $\mathcal{A}_t$  into  $\mathcal{B}(\mathcal{H}_t)$  and  $F(t)$  is a projection it follows that  $\|j_t(X)\| \leq \|X\|$  and  $j_t(I_t) = F(t)$ . To check that  $j_t(X)j_t(Y) = j_t(XY)$  it is



from (2.7). Since  $j_t^0(X)F(t) = F(t)j_t^0(X)F(t)$  it follows that  $j_t(X)^* = j_t(X^*)$ .

To prove (ii) it is enough to check that, for  $s < t$ ,

$$\begin{aligned} & \langle \lambda((X_s, X_{s_2}, \dots, X_{s_m}), u), j_t^0(X) \lambda((Y_s, Y_{s_2}, \dots, Y_{s_m}), v) \rangle = \\ & \langle \lambda((X_s, X_{s_2}, \dots, X_{s_m}), u), \lambda((T(s, t)(X) Y_s, Y_{s_2}, \dots, Y_{s_m}), v) \rangle \end{aligned}$$

for all  $X \in \mathcal{A}_t$ . By definitions the left hand side is equal to

$$\langle \lambda((I_t, X_s, X_{s_2}, \dots, X_{s_m}), u), \lambda((X, Y_s, Y_{s_2}, \dots, Y_{s_m}), v) \rangle$$

which, by property (i) in Proposition 2.2 and 2.3, is equal to the right hand side.

(iii) is just a restatement of property (ii) in Proposition 2.2 because

$$j_{t_1}(X_1) \dots j_{t_n}(X_n) V u = \lambda(X(\sigma), u)$$

with  $\sigma = \{t_1, t_2, \dots, t_n\}$ .

The first part of (iv) is contained in the last part of Proposition 2.4. The remaining part of (iv) follows from property (i) in Proposition 2.2. ■

*Remark.* It is interesting to compare the properties of  $\{F(t)\}$  and  $\{j_t\}$  in Theorem 2.5 with (1.2)–(1.4) in the case of classical Markov processes. This motivates the following definition: suppose  $\mathcal{A}_t, \mathcal{X}_t$  and  $T(s, t), s \leq t$  are as in Theorem 2.5. Then any quadruple  $(\mathcal{H}, F, \{j_t\}, V)$  consisting of a Hilbert space  $\mathcal{H}$ , an increasing family  $\{F(t)\}$  of projections in  $\mathcal{H}$ , contractive \* homomorphisms  $j_t$  from  $\mathcal{A}_t$  into  $\mathcal{B}(\mathcal{H})$  and a unitary isomorphism  $V$  from  $\mathcal{X}_0$  onto the range of  $F(0)$  is called a *conservative Markov flow* with transition operators  $T(\cdot, \cdot)$  if

$$j_t(I_t) = F(t), \quad F(s)j_t(X)F(s) = j_s(T(s, t)(X)) \text{ for } 0 \leq s \leq t < \infty$$

and  $j_0(X) V = V X$  for all  $X \in \mathcal{A}_0$ , the flow is said to be *minimal* if, in addition, property (iii) of Theorem 2.5 holds. Two such minimal conservative Markov flows  $(\mathcal{H}, F, \{j_t\}, V)$  and  $(\mathcal{H}', F', \{j'_t\}, V')$  with the same transition operators  $T(\cdot, \cdot)$  are called *equivalent* if there exists a unitary isomorphism  $W: \mathcal{H} \rightarrow \mathcal{H}'$  such that

$$W F(t) W^{-1} = F'(t), \quad W j_t(X) W^{-1} = j'_t(X), \quad W V = V'$$

for all  $t \geq 0, X \in \mathcal{A}_t$  [BP], [M].

We shall establish soon that upto equivalence the minimal Markov flow constructed in Theorem 2.5 is unique.

## PROPOSITION 2.6.

Let  $(\mathcal{H}, F, \{j_t\}, V)$  be a minimal conservative Markov flow with transition operators  $T(\cdot, \cdot)$  then the following hold:

(i) Let  $0 \leq t_1 < t_2 > t_3 < \infty$ . Then for any  $X_i \in \mathcal{A}_{t_i}, i = 1, 2, 3$

$$j_{t_1}(X_1) j_{t_2}(X_2) j_{t_3}(X_3) = \begin{cases} j_{t_1}(X_1) T(t_1, t_2)(X_2) j_{t_3}(X_3) & \text{if } t_1 \geq t_3 \\ j_{t_1}(X_1) j_{t_3}(T(t_3, t_2)(X_2) X_3) & \text{if } t_1 < t_3 \end{cases}$$

(ii) Let  $\mathcal{N}$  be the set of all pairs of sequences of the form  $(t_1, t_2, \dots, t_n; X_1, X_2, \dots, X_n)$  where  $0 \leq t_1, t_2, \dots, t_n < \infty$ ,  $X_i \in \mathcal{A}_{t_i}$ ,  $i = 1, 2, \dots, n$ ,  $n = 1, 2, \dots$ . Then there exists a map  $\alpha: \mathcal{N} \rightarrow \mathcal{A}_0$  independent of the Markov flow such that

$$F(0)j_{t_1}(X_1)j_{t_2}(X_2)\cdots j_{t_n}(X_n)F(0) = j_0(\alpha(t, \mathbf{X})) \quad (2.8)$$

for all  $(t, \mathbf{X}) = (t_1, t_2, \dots, t_n; X_1, X_2, \dots, X_n) \in \mathcal{N}$ .

*Proof.* Let  $t_1, t_2, t_3$  be as in (i) and  $t_1 \geq t_3$ . Then

$$\begin{aligned} & j_{t_1}(X_1)j_{t_2}(X_2)j_{t_3}(X_3) \\ &= j_{t_1}(X_1)F(t_1)j_{t_2}(X_2)F(t_1)j_{t_3}(X_3) \\ &= j_{t_1}(X_1)j_{t_1}(T(t_1, t_2)(X_2))j_{t_3}(X_3) \\ &= j_{t_1}(X_1)T(t_1, t_2)(X_2)j_{t_3}(X_3), \end{aligned}$$

which proves the first part of (i). Its second part is proved in the same manner.

To prove (ii) observe that

$$\begin{aligned} & F(0)j_{t_1}(X_1)j_{t_2}(X_2)\cdots j_{t_n}(X_n)F(0) \\ &= j_0(I_0)j_{t_1}(X_1)j_{t_2}(X_2)\cdots j_{t_n}(X_n)j_0(I_0). \end{aligned} \quad (2.9)$$

Without loss of generality assume that  $0 < t_1 < t_2 < \dots < t_{k-1} > t_k$ . Then by (i) the product  $j_{t_{k-2}}(X_{k-2})j_{t_{k-1}}(X_{k-1})j_{t_k}(X_k)$  can be reduced to a product of size 2 of the form  $j_{t_{k-2}}(X'_{k-2})j_{t_k}(X_k)$  or  $j_{t_{k-2}}(X_{k-2})j_{t_k}(X'_k)$  where the primed operators depend only on  $(t, \mathbf{X})$  and  $T(\cdot, \cdot)$  and not on the particular flow under consideration. Thus the  $n$ -fold product between the two  $j_0(I_0)$ 's on the right hand side of (2.9) can be reduced to an  $(n-1)$ -fold product. A successive reduction of the sequence  $(0, t_1, t_2, \dots, t_n, 0; I_0, X_1, X_2, \dots, X_n, I_0)$  applying (i) yields in the end an element  $\alpha(t, \mathbf{X})$  satisfying (2.8). ■

**Theorem 2.7.** Let  $\mathcal{A}, \mathcal{H}, T(s, t)$ ,  $0 \leq s \leq t < \infty$  be as in Theorem 2.5. Then any two minimal conservative Markov flows with transition operators  $T(\cdot, \cdot)$  are equivalent.

*Proof.* Let  $(\mathcal{H}, F, \{j_i\}, V)$  and  $(\mathcal{H}', F', \{j'_i\}, V')$  be two Markov flows satisfying the conditions of the theorem. Suppose that  $s_1 > s_2 > \dots > s_m = 0$ ,  $t_1 > t_2 > \dots > t_n = 0$ ,  $X_i \in \mathcal{A}_{s_i}$ ,  $Y_j \in \mathcal{A}_{t_j}$ ,  $i = 1, 2, \dots, m$ ,  $j = 1, 2, \dots, n$ . Consider  $(\mathbf{r}, \mathbf{Z}) \in \mathcal{N}$  (where  $\mathcal{N}$  is as in Proposition 2.6) defined by

$$\begin{aligned} \mathbf{r} &= (s_m, s_{m-1}, \dots, s_1, t_1, t_2, \dots, t_n), \\ \mathbf{Z} &= (X_m^*, X_{m-1}^*, \dots, X_1^*, Y_1, Y_2, \dots, Y_n). \end{aligned}$$

Since  $s_m = t_n = 0$  it follows from Proposition 2.6 that there exists  $\alpha(\mathbf{r}, \mathbf{Z}) \in \mathcal{A}_0$  such that

$$\begin{aligned} & j_{s_m}(X_m^*)j_{s_{m-1}}(X_{m-1}^*)\cdots j_{s_1}(X_1^*)j_{t_1}(Y_1)\cdots j_{t_n}(Y_n) = j_0(\alpha(\mathbf{r}, \mathbf{Z})), \\ & j'_{s_m}(X_m^*)j'_{s_{m-1}}(X_{m-1}^*)\cdots j'_{s_1}(X_1^*)j'_{t_1}(Y_1)\cdots j'_{t_n}(Y_n) = j'_0(\alpha(\mathbf{r}, \mathbf{Z})). \end{aligned}$$

Thus for any  $u, v \in \mathcal{K}_0$  we have

$$\begin{aligned} & \langle j_{s_1}(X_1) \cdots j_{s_m}(X_m) Vu, j_{t_1}(Y_1) \cdots j_{t_n}(Y_n) Vv \rangle \\ & \quad \langle j'_{s_1}(X_1) \cdots j'_{s_m}(X'_m) V'u, j'_{t_1}(Y_1) \cdots j'_{t_n}(Y_n) V'v \rangle \\ & = \langle u, \alpha(\mathbf{r}, \mathbf{Z})v \rangle. \end{aligned}$$

From the minimality of the two flows it follows that  $\mathcal{H}$  and  $\mathcal{H}'$  are spanned by vectors of the form  $j_{t_1}(Y_1) \cdots j_{t_n}(Y_n) Vu$  and  $j'_{t_1}(Y_1) \cdots j'_{t_n}(Y_n) V'u$  respectively. Hence there exists a unitary isomorphism  $W: \mathcal{H} \rightarrow \mathcal{H}'$  such that

$$Wj_{t_1}(Y_1) \cdots j_{t_n}(Y_n) Vu = j'_{t_1}(Y_1) \cdots j'_{t_n}(Y_n) V'u$$

for all  $u \in \mathcal{K}_0$ ,  $t_1 > t_2 > \cdots > t_n = 0$ ,  $Y_i \in \mathcal{A}_{t_i}$ ,  $i = 1, 2, \dots, n$ . That  $W$  is the required isomorphism implementing the equivalence of the two flows is immediate. ■

*Remark.* Let  $(\mathcal{H}, F, \{j_t\}, V)$  be a minimal conservative Markov flow with transition operators  $T(\cdot, \cdot)$ . Denote by  $\mathcal{B}$  and  $\mathcal{B}_t$  respectively the  $C^*$  algebras generated by  $\{j_s(X), X \in \mathcal{A}_s, 0 \leq s < \infty\}$  and  $\{j_s(X), X \in \mathcal{A}_s, 0 \leq s \leq t\}$ . By the same arguments as in the proof of Proposition 2.6 it is easy to see that for  $t_i \geq s$ ,  $i = 1, 2, \dots, n$  an expression of the form  $F(s)j_{t_1}(X_1) \cdots j_{t_n}(X_n)F(s)$  can be expressed as  $j_s(\alpha_s(\mathbf{t}, \mathbf{X}))$  where  $\alpha_s(\mathbf{t}, \mathbf{X}) \in \mathcal{A}_s$ . In particular the map  $\mathbb{E}_s$  defined by

$$\mathbb{E}_s(Z) = F(s)ZF(s), \quad Z \in \mathcal{B}$$

maps  $\mathcal{B}$  onto  $\mathcal{B}_s$ . We may call  $\mathbb{E}_s$  the *conditional expectation map* from  $\mathcal{B}$  onto  $\mathcal{B}_s$ . If  $\rho_0$  is a state on  $\mathcal{A}_0$  then a state  $\rho$  on  $\mathcal{B}$  is uniquely determined by

$$\rho(Z) = \rho_0(V^*F(0)ZF(0)V), \quad Z \in \mathcal{B}.$$

It is legitimate to call the filtered quantum probability space  $(\mathcal{B}, \mathcal{B}_t, \rho)$  the Markov process with initial state  $\rho_0$  and transition operators  $T(\cdot, \cdot)$ .

Let  $\mathcal{Z}_t$  denote the centre of  $\mathcal{A}_t$  for each  $t$ . It is possible that  $T(s, t)$  may not map  $\mathcal{Z}_t$  into  $\mathcal{Z}_s$ . In the minimal flow with transition operators  $T(\cdot, \cdot)$ , the operators  $\{j_t(Z), Z \in \mathcal{Z}_t, t \geq 0\}$  need not be a commutative family. However, by following an idea in Bhat [B], we shall modify the construction in Proposition 2.4 in order to arrive at a family of  $*$  unital homomorphisms  $k_t: \mathcal{Z}_t \rightarrow \mathcal{B}(\mathcal{H})$  so that  $\{k_t(Z), Z \in \mathcal{Z}_t, t \geq 0\}$  is a commutative family and  $j_t(Z)$  is obtained from  $k_t(Z)$  by a conditional expectation.

**Theorem 2.8.** *Let  $(\mathcal{H}, F, \{j_t\}, V)$  be as in Theorem 2.5. Then there exists a unique  $*$ unital homomorphism  $k_t: \mathcal{Z}_t \rightarrow \mathcal{B}(\mathcal{H})$  satisfying the following:*

(i) for any  $t_1 > t_2 > \cdots > t_n = 0$ ,  $X_i \in \mathcal{A}_{t_i}$ ,  $i = 1, 2, \dots, n$ ,  $Z \in \mathcal{Z}_t$  and  $u \in \mathcal{K}_0$

$$k_t(Z)\lambda((X_{t_1}, X_{t_2}, \dots, X_{t_n}), u) = \begin{cases} \lambda((X_{t_1}, X_{t_2}, \dots, X_{t_{i-1}}, ZX_{t_i}, X_{t_{i+1}}, \dots, X_{t_n}), u) \\ \quad \text{if } t = t_i \text{ for some } i \\ \lambda((Z, X_{t_1}, \dots, X_{t_n}), u) \quad \text{if } t > t_1, \\ \lambda((X_{t_1}, X_{t_2}, \dots, X_{t_{i-1}}, ZX_{t_i}, \dots, X_{t_n}), u) \\ \quad \text{if } t_{i-1} > t > t_i \text{ for some } i; \end{cases} \quad (2.10)$$

(ii) the family  $\{k_t(Z), Z \in \mathcal{Z}_t, t \geq 0\}$  is commutative;

(iii)  $j_t(Z) = F(t)k_t(Z)F(t)$  for all  $t \geq 0$ ,  $Z \in \mathcal{Z}_t$ .

*Proof.* As in the proof of Proposition 2.4 consider a unitary element  $Z \in \mathcal{Z}_t$ . Suppose  $t = t_i$  for some  $i = 1, 2, \dots, n$ . For any  $X_{t_i}, Y_{t_i} \in \mathcal{Z}_{t_i}, i = 1, 2, \dots, n$  we have

$$\begin{aligned} & \langle \lambda((X_{t_1}, X_{t_2}, \dots, X_{t_{i-1}}, ZX_{t_i}, X_{t_{i+1}}, \dots, X_{t_n}), u), \\ & \lambda((Y_{t_1}, Y_{t_2}, \dots, Y_{t_{i-1}}, ZY_{t_i}, Y_{t_{i+1}}, \dots, Y_{t_n}), v) \rangle = \\ & \langle u, X_{t_n}^* (\dots X_{t_i}^* Z^* T(t_i, t_{i-1}) (\dots (X_{t_2}^* T(t_2, t_1) (X_{t_1}^* Y_{t_1}) Y_{t_2}) \dots) ZY_{t_i} \dots) Y_{t_n} v \rangle. \end{aligned}$$

Since  $Z$  and  $Z^* \in \mathcal{Z}_{t_i}$  and  $Z^*Z = 1$  it follows that the right hand side is independent of  $Z$ . The same argument in the remaining cases together with the Chapman-Kolmogorov equations for  $T(\cdot, \cdot)$  and (iv) in Proposition 2.2 imply that  $k_t(Z)$  defined by (2.10) on elements of the form  $\lambda(X(\sigma), u)$  is scalar product preserving. Hence  $k_t(Z)$  extends to a unitary operator on  $\mathcal{H}$ . Furthermore for any two unitary elements  $Z, Z' \in \mathcal{Z}_t$ , we have  $k_t(Z)k_t(Z') = k_t(ZZ')$ . Once again by (iv) in Proposition 2.2,  $k_t(I_t)$  is the identity operator in  $\mathcal{H}$ . Exactly as in the proof of Proposition 2.4 we extend  $k_t(\cdot)$  to a  $*$  unital homomorphism from  $\mathcal{Z}_t$  into  $\mathcal{B}(\mathcal{H})$ . This proves (i).

If  $t \neq t'$ ,  $Z \in \mathcal{Z}_t$ ,  $Z' \in \mathcal{Z}_{t'}$ , it follows from (2.10) by straightforward verification that

$$k_t(Z)k_{t'}(Z')\lambda(X(\sigma), u) = k_{t'}(Z')k_t(Z)\lambda(X(\sigma), u)$$

where  $\sigma = \{t_1 > t_2 > \dots > t_n = 0\}$ . This proves (ii).

When  $t = t_1 > t_2 > \dots > t_n$ ,  $X_{t_i}, Y_{t_i} \in \mathcal{Z}_{t_i}$ ,  $u, v \in \mathcal{H}_0$  we have

$$\begin{aligned} & \langle \lambda((X_1, X_{t_2}, \dots, X_{t_n}), u), k_t(Z)\lambda((Y_1, Y_{t_2}, \dots, Y_{t_n}), v) \rangle \\ & = \langle \lambda((X_1, X_{t_2}, \dots, X_{t_n}), u), \lambda((ZY_1, Y_{t_2}, \dots, Y_{t_n}), v) \rangle \\ & = \langle \lambda((X_1, X_{t_2}, \dots, X_{t_n}), u), j_t(Z)\lambda((Y_1, Y_{t_2}, \dots, Y_{t_n}), v) \rangle. \end{aligned}$$

Since vectors of the form  $\lambda((X_1, X_{t_2}, \dots, X_{t_n}), u)$  span the range  $\mathcal{H}_t$  of  $F(t)$ , property (iii) is immediate. Uniqueness of  $\{k_t\}$  follows from the minimality of  $\{j_t\}$  and property (i). ■

## References

- [AFL] Accardi L, Frigerio A and Lewis J T, Quantum stochastic processes, *Publ. RIMS, Kyoto Univ.* **18** (1982) 97-133
- [B] Bhat B V, Rajarama, *Markov dilations of nonconservative quantum dynamical semigroups and a quantum boundary theory*, Ph.D. Thesis (submitted to Indian Statistical Institute, New Delhi) (1993)
- [BP] Bhat B V, Rajarama and Parthasarathy K R, *Markov dilations of nonconservative quantum dynamical semigroups and a quantum boundary theory*, Indian Statistical Institute, New Delhi preprint (1993)
- [E] Emch G G, Minimal dilations of CP flows, *C\* algebras and applications to physics*, Springer LNM **650** (1978) 156-159
- [M] Meyer P A, Processus de Markov non-commutatifs d'après Bhat-Parthasarathy (to appear in *Séminaire de Probabilités* (ed.) Meyer Azéma and Yor) (Strasbourg preprint) (1993)
- [P] Parthasarathy K R, An Introduction to Quantum Stochastic Calculus, *Monographs in Mathematics* (Basel: Birkhauser Verlag) (1992)
- [S] Sauvageot J L, Markov quantum semigroups admit covariant Markov  $C^*$  dilations, *Commun. Math. Phys.* **106** (1986) 91-103
- [Vi-S] Vincent-Smith G F, Dilations of a dissipative quantum dynamical system to a quantum Markov process, *Proc. Lond. Math. Soc.* (3) **49** (1984) 58-72

## Iterations of random and deterministic functions

K B ATHREYA

Department of Mathematics, 400 Carver Hall, Iowa State University, Ames, IA 50010–2066, USA

Dedicated to the memory of Professor K G Ramanathan

**Abstract.** Let  $f$  be a probability generating function on  $[0, 1]$ . The convergence of its iterates  $f_n$  to fixed points is studied in this paper. Results include rates for  $f$  and  $f^{-1}$ . Also iterates of independent identically distributed stable processes are studied and a trichotomy based on the order of the stability is established.

**Keywords.** Random maps; generating functions; iteration; stable processes; auto regressive processes.

### 1. Introduction

Let  $f$  be map from an interval  $I$  on the real line into itself. Let  $f_0(s) \equiv s, f_{n+1}(s) = f(f_n(s))$  for  $n \geq 0$ . The sequence  $\{f_n(s)\}_0^\infty$ , called *the iterates of  $f$* , arise in many areas of mathematics. The problems that are studied in this connection include finding all the fixed points, the rates of convergence to their fixed points and similar aspects for the iterates of the inverse  $f^{-1}$  of  $f$ . In this paper we study this for a class of  $f$ 's that are generating functions of probability distributions on the nonnegative integers. These are motivated by applications to the theory of branching processes (see [2]).

Let  $X_0(t, \tilde{\omega})$  be map from  $T \times \tilde{\Omega}$  to  $T$  where  $T$  and  $\tilde{\Omega}$  are nonempty sets. Let  $\tilde{B}$  be a  $\sigma$ -algebra of subsets of  $\tilde{\Omega}$  and  $\tilde{P}$  be a probability measure on  $\tilde{B}$ . Then under appropriate measurability conditions,  $\{X_0(t, \tilde{\omega}): t \in T\}$  is called a  $T$ -valued stochastic process on  $(\tilde{\Omega}, \tilde{B}, \tilde{P})$  with index set  $T$ . Let  $(\Omega, B, P)$  be a probability space on which is defined a sequence  $\{X_i(t, \omega)\}_{i=1, 2, \dots}$  of independent identically distributed copies of the stochastic process  $X_0$ . Let  $Y_0(t, \omega) \equiv t$  and  $Y_{n+1}(t, \omega) = X_{n+1}(Y_n(t, \omega), \omega)$  for  $n = 0, 1, 2, \dots$ . The sequence  $\{Y_n(t, \omega)\}_0^\infty$ , called *the i.i.d iterates of the stochastic process  $X_0$* , arise in many areas of probability theory. The problems studied include convergence of these iterates and rates of convergence. In this paper we study the case when  $T = (-\infty, \infty)$  and the process  $X$  has a self similarity property of the form  $X_0(t, \tilde{\omega})$  has the same distribution as  $|t|^{1/\alpha} X_0(1, \tilde{\omega})$ . These include Brownian motion and stable processes. We also study the case when  $X_0$  is a random walk on the integers. Finally we indicate some open problems.

### 2. Iterates of probability generating functions

Let  $\{p_j\}_0^\infty$  be a sequence of numbers satisfying  $p_j \geq 0, \sum_0^\infty p_j = 1$ . Then for  $s$  real,

$$f(s) \equiv \sum_0^{\infty} p_j s^j \quad (1)$$

is called the (probability) generating function of the sequence  $\{p_j\}$ . It is convergent for  $|s| \leq 1$ . The iterates  $\{f_n(s)\}_{n=0}^{\infty}$  have a nice interpretation in terms of a class of stochastic processes known as *branching processes*. Let  $T = N^+ \equiv \{0, 1, 2, \dots\}$  and  $X_0(t, \tilde{\omega})$  be a random walk with step distribution  $\{p_j\}$ . That is,  $X_0(0, \tilde{\omega}) = 0$ ,  $X_0(n, \tilde{\omega}) = \sum_{j=1}^n \zeta_j(\tilde{\omega})$  for  $n \geq 1$  where  $\{\zeta_i(\tilde{\omega})\}_1^{\infty}$  are independent random variables with distribution  $\{p_j\}_0^{\infty}$ . Let  $\{Y_n(t, \omega)\}_0^{\infty}$  be the i.i.d. iterate sequence of  $X_0$  as defined in the introduction. Then using  $EZ$  to denote the expectation, i.e. the integral of a  $B$ -measurable function  $Z(\omega)$  with respect to the measure  $P$  on the space  $(\Omega, B, P)$ , it can be verified that for  $0 \leq s \leq 1$

$$E(s^{Y_n(1, \omega)}) = f_n(s) \text{ for } n \geq 1. \quad (2)$$

Indeed, it is not difficult to show that the conditional expectation

$$E(s^{Y_{n+1}(1, \omega)} | Y_n(1, \omega), Y_{n-1}(1, \omega), \dots, Y_0(1, \omega)) \quad (3)$$

is  $f(s)^{Y_n(1, \omega)}$  (using the independence of  $X_{n+1}$  and  $Y_n$  and the random walk nature of  $X_{n+1}$ ). This, in turn, implies (2). The sequence  $Z_n \equiv Y_n(1, \omega)$   $n = 0, 1, 2, \dots$  is called a Galton-Watson branching process (see [2]). Thus, there is a connection between the deterministic iterates sequence  $\{f_n(s)\}_0^{\infty}$  and the stochastic iterates sequence  $\{Y_n\}_0^{\infty}$ . It can be shown that  $f(s)$  has almost two fixed points in  $[0, 1]$  and  $f_n(s)$  converges to one of these as  $n \rightarrow \infty$ . This in turn is related to the instability of the sequence  $\{Z_n = Y_n(1, \omega)\}$  in the sense that  $Z_n$  either goes to  $\infty$  or to zero as  $n \rightarrow \infty$ .

The following result is known. See [2], Chapter 1.

**Theorem 1.** a) There are at most two solutions to  $f(s) = s$  in  $[0, 1]$ . b) If  $m = \sum j p_j > 1$  there are exactly two solutions,  $q$  and 1, with  $0 \leq q < 1$ . c) If  $m \leq 1$  then there is only one solution, namely,  $q = 1$ . d) For  $0 \leq s < 1$ ,  $f_n(s) \rightarrow q$  where  $q$  is the smaller of the two roots of  $f(s) = s$  in  $[0, 1]$ . e) For any distribution of the initial value  $Y_0(t, \omega)$

$$P(Y_n(t, \omega) \rightarrow 0) + P(Y_n(t, \omega) \rightarrow \infty) = 1. \quad (4)$$

An interesting question that arises from part d of Theorem 1 is about the rate of convergence of  $f_n(s)$  to  $q$  (See [2] Chapter 1).

**Theorem 2.** Let  $m = \sum j p_j \neq 1$  and  $\gamma = f'(q) \neq 0$ . Then,

$$\lim_{n \rightarrow \infty} \frac{f_n(s) - q}{\gamma^n} \equiv Q(s) \quad (5)$$

exists for  $0 \leq s < 1$  and  $Q(\cdot)$  is the unique solution of the functional equation

$$Q((s) = \gamma Q(s)) \text{ for } 0 \leq s < \max(q, 1) \quad (6)$$

subject to  $Q(q) = 0$ . Further, if  $m > 1$  then  $\lim_{s \uparrow 1} Q(s) = 0$ .

An application of Theorem 2 yields the following large deviation result for branching processes.

**Theorem 3.** Assume  $p_0 = 0, p_1 > 0$  and  $\sum j^{2r+\delta} p_j < \infty$  for some  $r \geq 1$  and  $\delta > 0$  such that  $p_1 m^r > 1$ . Then

$$\lim_{n \rightarrow \infty} \frac{1}{p_1^n} P\left(\left|\frac{Z_{n+1}}{Z_n} - m\right| > \varepsilon\right) = \sum_k q_k \phi(k, \varepsilon) < \infty, \quad (7)$$

where  $\{q_k\}$  is the sequence with generating function  $Q(s)$  as in (5) and (6) and

$$Z_n \equiv Y_n(1, \omega), \phi(k, \varepsilon) = P\left(\left|\frac{Y_1(k, \omega)}{k} - m\right| > \varepsilon\right). \quad (8)$$

There is a counterpart to Theorem 2 when  $p_1 = 0 = p_0$ .

**Theorem 4.** Let  $p_0 = 0 = p_1$  and  $k = \inf\{j: j \geq 1, p_j \neq 0\}$ . Then

$$f_n(s) = s^{k^n} p_k^{\sum_{j=0}^{n-1} k^j} (R_n(s))^{k^n}, \quad (9)$$

where  $\lim_n R_n(s) = R(s)$  exists uniformly in  $[0, 1]$  with  $R(0) = 1$  and  $R(1) < \infty$ . Further,

$$(f_n(s))^{1/k^n} \rightarrow p_k^{1/(k-1)} s R(s) \quad (10)$$

An application of Theorem 4 is the following

**Theorem 5.** Let  $p_0 = 0 = p_1$  and  $k = \inf\{j: j \geq 1, p_j \neq 0\}$ . Let  $f(s_0) = \sum s_0^j p_j < \infty$  for some  $1 < s_0 < \infty$ . Then for each  $\varepsilon > 0$  there exists  $C_\varepsilon$  and  $\lambda_\varepsilon$  such that  $0 < \lambda_\varepsilon < 1$  and

$$P\left(\left|\frac{Z_{n+1}}{Z_n} - m\right| > \varepsilon\right) \leq C_\varepsilon \lambda_\varepsilon^{k^n}. \quad (11)$$

**Theorem 6.** Let  $p_0 = 0$  and  $g = f^{-1}$  defined by  $f(g(s)) = s$ . Let  $f(s_0) = \sum_j p_j s_0^j$  be  $< \infty$  for some  $s_0 > 1$ . Then for  $1 \leq s \leq f(s_0), g_n(s) \downarrow 1$  and

$$\tilde{Q}_n(s) \equiv m^n (q_n(s) - 1) \downarrow \tilde{Q}(s). \quad (12)$$

A consequence of Theorem 6 is the following

**Theorem 7.** Let  $f(s_0) = \sum p_j s_0^j$  be  $< \infty$  for some  $s_0 > 1$ . Then there exists a  $\theta_1 > 0$  such that

$$\sup_n E(\exp(\theta_1 W_n)) < \infty, \quad (13)$$

where  $W_n = Z_n m^{-n}$ . Further there exist constants  $C$  and  $\lambda > 0$  such that for each  $\varepsilon > 0$

$$P(|W_n - W| \geq \varepsilon) \leq C \exp\left(-\left(\lambda \varepsilon \frac{2}{3} \left(m \frac{1}{3}\right)^n\right)\right), \quad (14)$$

where  $W = \lim_n W_n$ .

The proofs of Theorem 3-7 are in Athreya [1].

The extensions of above results to generating functions of probability measures on the nonnegative lattices in Euclidean spaces are contained in the thesis of Vidyashankar [4].

### 3. Iterations of random processes

Let  $\{X_i(t, \omega); t \in T\}_1^\infty$  be a sequence of  $T$ -valued stochastic processes with index set  $T$  and defined on a probability space  $(\Omega, B, P)$ . Let

$$Y_0(t, \omega) = t, Y_{n+1}(t, \omega) = X_{n+1}(Y_n(t, \omega)) \text{ for } n = 0, 1, 2, \dots \quad (15)$$

We study the iterate sequence  $\{Y_n\}$  when  $\{X_i\}$  are independent and identically distributed copies of a stable process of order  $\alpha$  on  $R$ . Let  $\{X(t, \omega); t \in R\}$  be a real valued stochastic process such that:

- i)  $X(0, \omega) = 0$  w.p.1. and
- ii) Both  $\{X(t, \omega); t \geq 0\}$  and  $\{X(-t, \omega); t \geq 0\}$  be independent copies of a real valued stochastic process with independent and stationary increments.

Let  $\{X_n(t, \omega); -\infty < t < \infty\}_1^\infty$  be independent copies of  $X$ . Let  $\{Y_n(t, \omega)\}_1^\infty$  be as in (15).

Let  $\mu(t, \cdot)$  be the probability distribution of  $X(t, \omega)$  for  $t \geq 0$ . Then for each fixed  $t$ ,  $\{Y_n(t, \omega)\}_n = 0, 1, \dots$  is a Markov chain with state space  $R$ , stationary transition probability function  $P(x, A) = \mu(|x|, A)$  for  $x \in R, A \in B(R)$  the Borel  $\sigma$ -algebra of  $R$ .

**Theorem 8.** Let  $\mu(t, A) = \Phi\left(\frac{A}{\sqrt{t}}\right)$  where

$$\Phi(A) = \int_A \frac{1}{\sqrt{2\pi}} \exp - \left(\frac{x^2}{2}\right) dx \text{ for } A \in B(R).$$

Then for any  $t \neq 0$ ,  $Y_n(t, \omega)$  converges in distribution and the limit is independent of  $t$ .

More generally let  $\mu_\alpha(t, \cdot)$  be the distribution of a symmetric stable process of order  $\alpha$ .

**Theorem 9.** If  $\mu(|t|, \cdot) = \mu_\alpha(t, \cdot)$  and  $1 \leq \alpha \leq 2$  then for any  $t \neq 0$ ,  $Y_n(t, \omega)$  converges in distribution and the limits is independent of  $t$  and its distribution function is given by

$$F(y) = \int_0^\infty G(y|x|^{-\alpha}) dH(x) \text{ where } G(x) = P(X_1(1) \leq x)$$

**Theorem 10.** If  $\mu(|t|, \cdot) = \mu_\alpha(t, \cdot)$  and  $0 < \alpha < 1$ . Then for any  $t \neq 0$ ,  $\lim_n |Y_n(t, \omega)|^{\alpha^n}$  exists w.p. 1 and equals  $\exp(|t| + Z(t, \omega))$  where  $Z(t, \omega)$  is a random variable with distribution that is identical to that of  $\sum_{j=1}^\infty \alpha^j \eta_j$  where  $\{\eta_j, j = 1, 2, \dots\}$  are independent with distribution same as that of  $\log|X(1)|$ .



**Theorem 11.** If  $\mu(t, \cdot) = \mu_1(|t|, \cdot)$  (i.e.  $\alpha = 1$ ) then

$$\lim_n |Y_n(t, \omega)| = \begin{cases} \infty & \text{w.p.1 if } \mu = E \log(X, (1)) > 0 \\ 0 & \text{w.p.1} < 0 \end{cases}$$

and if  $\mu = 0$  then w.p.1 both 0 and  $\infty$  are limit points for the sequence  $\{|Y_n(t, \omega)|\}$ . Also  $(|Y_n| \exp - (n\mu))^{1/\sqrt{n}}$  converges in distribution to  $\exp(N\sigma)$  where  $N \sim N(0, 1)$  and  $\sigma^2 = \text{variance of } \log|X_1(1)|$ .

The proofs of Theorems 8–11 depend on the following representation. From the definition of  $Y_n$  it follows that

$$|Y_{n+1}(t, \omega)| = \frac{|X_{n+1}(Y_n(t, \omega))|}{|Y_n(t, \omega)|^{1/\alpha}} |Y_n(t, \omega)|^{1/\alpha}.$$

Taking logarithms we see that  $|Y_n(t, \omega)| \equiv Z_n$  is an autoregressive sequence satisfying

$$Z_{n+1} = \rho Z_n + \eta_{n+1}$$

where  $\rho = (1/\alpha)$ ,  $\{\eta_n\}$  are i.i.d with distribution same as  $\log|X_1(1)|$ . For  $1 < \alpha \leq 2$ ,  $\frac{1}{2} \leq \rho < 1$  and hence  $\{Z_n\}$  is nonexplosive. For  $0 < \alpha < 1$ ,  $\rho > 1$  and so  $\{Z_n\}$  is explosive. Details of the proofs of Theorems 8–11 will appear in a future publication.

#### 4. Some open problems

- In the stable processes case do the joint distributions of  $(Y_n(t_i, \omega), i = 1, 2, \dots, k)$  converge as  $n \rightarrow \infty$  where  $t_1, t_2, \dots, t_k$  are fixed in  $t$ ?
- What happens in the case when the stable processes are replaced by a general additive process with stationary independent increments? Are there reasonable conditions in terms of the Levy measure for convergence? When  $T$  is the set of integers and the underlying  $X$  process is a random walk the  $\{Y_n\}$  sequence yields the self annihilating branching studied by Erickson [3]. It would be worthwhile to extend his results to the Levy process case.
- Suppose that the  $\{X_i\}$  are Markov chains that are positive recurrent. This corresponds to random dynamical systems that are Markov chains in random environments. What happens to  $\{Y_n\}$  in this case?

#### Acknowledgement

Research supported in part by NSF grant DMS 9343932.

#### References

- [1] Athreya K. B, Large deviation rates for branching processes – I single type case (1993), Preprint, Dept. of Mathematics, Iowa State University (To appear in the *Ann. Appl. Probab.*)
- [2] Athreya K. B and Ney P, Branching processes (1972) (New York: Springer-Verlag)
- [3] Erickson K. B, Self annihilating branching processes, *Ann. Probab.* 1 (1973) 926–946
- [4] Vidyashankar A, Large deviation rates for branching processes (1994), Ph.D. Thesis, Dept. of Mathematics, Iowa State University.



## Existence theory for linearly elastic shells

PHILIPPE G CIARLET

Laboratoire d'Analyse Numerique, Tour 55 Université Pierre et Marie Curie, 4 Place Jussieu,  
75005 Paris, France

Dedicated to the memory of Professor K G Ramanathan

**Abstract.** We review existence and uniqueness results, recently obtained for three of the most important linear two-dimensional shell models: Koiter's model, the bending model and the membrane model. They rely on a crucial lemma of J L Lions, used in an essential way for establishing in each case a generalized Korn's inequality, which is then combined with a generalized rigid displacement lemma of a geometrical nature.

**Keywords.** Linearly elastic shells; bending shell model; membrane shell model; Koiter's model.

### 1. Geometrical and mechanical preliminaries

In what follows, Greek indices and exponents vary in the set  $\{1, 2\}$ , Latin indices and exponents vary in the set  $\{1, 2, 3\}$ , and the repeated index or exponent convention for summation is used. The Euclidean inner product, the vector product and the Euclidean norm, of vectors  $\mathbf{u}, \mathbf{v} \in \mathbf{R}^3$  are denoted as  $\mathbf{u} \cdot \mathbf{v}$ ,  $\mathbf{u} \times \mathbf{v}$ , and  $|\mathbf{u}|$ .

Let  $\omega$  be an open, bounded, connected subset of  $\mathbf{R}^2$  with a Lipschitz-continuous boundary  $\gamma$ , the set  $\omega$  being locally on one side of  $\gamma$ . Let  $y = (y^1, y^2)$  denote a generic point of the set  $\bar{\omega}$  and let  $\partial_\alpha = \partial/\partial y^\alpha$ . We consider a surface  $S$  in  $\mathbf{R}^3$ , of the form  $S = \varphi(\bar{\omega})$ , where  $\varphi: \bar{\omega} \rightarrow \mathbf{R}^3$  is a given, injective, smooth enough mapping. We assume that the two vectors  $\mathbf{a}_\alpha = \partial_\alpha \varphi$  are linearly independent at all points of  $\bar{\omega}$ .

The vectors  $\mathbf{a}_\alpha$  form the covariant basis of the tangent plane, and the vectors  $\mathbf{a}^\alpha$ , defined by the relations  $\mathbf{a}^\alpha \cdot \mathbf{a}_\beta = \delta^\alpha_\beta$ , form its contravariant basis. The three vectors  $\mathbf{a}^i$ , where  $\mathbf{a}^3 = \mathbf{a}_3 = (\mathbf{a}_1 \times \mathbf{a}_2)/|\mathbf{a}_1 \times \mathbf{a}_2|$  form the contravariant basis at each point of  $S$ . The Christoffel symbols are defined by

$$\Gamma_{\alpha\beta}^\rho = \mathbf{a}^\rho \cdot \partial_\alpha \mathbf{a}_\beta,$$

and the first, second and third fundamental forms of  $S$  are defined by

$$a_{\alpha\beta} = \mathbf{a}_\alpha \cdot \mathbf{a}_\beta \text{ or } a^{\alpha\beta} = \mathbf{a}^\alpha \cdot \mathbf{a}^\beta,$$

$$b_{\alpha\beta} = -\mathbf{a}_\alpha \cdot \partial_\beta \mathbf{a}_3,$$

$$c_{\alpha\beta} = b_\alpha^\rho b_{\rho\beta}, \text{ where } b_\alpha^\rho = a^{\rho\sigma} b_{\sigma\alpha}.$$

Note that  $\Gamma_{\alpha\beta}^\rho = \Gamma_{\beta\alpha}^\rho$ ,  $a_{\alpha\beta} = a_{\beta\alpha}$ ,  $b_{\alpha\beta} = b_{\beta\alpha}$ ,  $c_{\alpha\beta} = c_{\beta\alpha}$ . Finally, we let

$$a = \det(a_{\alpha\beta}).$$

We consider a linearly elastic shell, with middle surface  $S$  and thickness  $2\varepsilon$ , clamped along a portion of its lateral face, and we let  $\lambda > 0$  and  $\mu > 0$  denote the Lamé constants of its constituting material. In each one of the two-dimensional shell models considered here, the unknowns are the three covariant components  $\zeta_i: \bar{\omega} \rightarrow \mathbf{R}$  of the displacement  $\zeta_i \mathbf{a}^i$  of the points of  $S$ , and we let  $\zeta = (\zeta_i): \bar{\omega} \rightarrow \mathbf{R}^3$ . With an arbitrary vector field  $\boldsymbol{\eta} = (\eta_i): \bar{\omega} \rightarrow \mathbf{R}^3$ , we associate the (linearized) strain, or change of metric, tensor and the (linearized) change of curvature tensor, whose covariant components are respectively given by

$$\begin{aligned}\gamma_{\alpha\beta}(\boldsymbol{\eta}) &= \frac{1}{2}(\partial_\alpha \eta_\beta + \partial_\beta \eta_\alpha) - \Gamma_{\alpha\beta}^\rho \eta_\rho - b_{\alpha\beta} \eta_3, \\ \Upsilon_{\alpha\beta}(\boldsymbol{\eta}) &= \partial_{\alpha\beta} \eta_3 - \Gamma_{\alpha\beta}^\rho \partial_\rho \eta_3 - c_{\alpha\beta} \eta_3 \\ &\quad + b_\beta^\rho (\partial_\alpha \eta_\rho - \Gamma_{\rho\alpha}^\sigma \eta_\sigma) + b_\alpha^\rho (\partial_\beta \eta_\rho - \Gamma_{\rho\beta}^\sigma \eta_\sigma) \\ &\quad + (\partial_\alpha b_\beta^\rho + \Gamma_{\alpha\sigma}^\rho b_\beta^\sigma - \Gamma_{\alpha\beta}^\sigma b_\sigma^\rho) \eta_\rho.\end{aligned}$$

We shall also use the fourth-order elasticity tensor of a two-dimensional shell, whose contravariant components are given by

$$a^{\alpha\beta\rho\sigma} = \frac{4\lambda\mu}{(\lambda + 2\mu)} a^{\alpha\beta} a^{\rho\sigma} + 2\mu(a^{\alpha\rho} a^{\beta\sigma} + a^{\alpha\sigma} a^{\beta\rho}).$$

## 2. The two-dimensional shell model of W T Koiter

The fundamental work of John [17] has led Koiter [18] to propose the following two-dimensional shell model, called Koiter's model: The unknown  $\zeta = (\zeta_i)$  solves the following variational problem:

$$\zeta \in \mathbf{V}(\omega) \text{ and } B(\zeta, \boldsymbol{\eta}) = L(\boldsymbol{\eta}) \text{ for all } \boldsymbol{\eta} \in \mathbf{V}(\omega),$$

where  $(\partial_\nu)$  denotes the outer normal derivative along  $\gamma$ , and  $\gamma_0$  is a subset of the boundary  $\gamma$ :

$$\begin{aligned}\mathbf{V}(\omega) &= \{\boldsymbol{\eta} = (\eta_i); \eta_\alpha \in H^1(\omega), \eta_3 \in H^2(\omega), \eta_i = \partial_\nu \eta_3 = 0 \text{ on } \gamma_0\}, \\ B(\zeta, \boldsymbol{\eta}) &= \int_\omega \left\{ \frac{\varepsilon^3}{3} a^{\alpha\beta\rho\sigma} \Upsilon_{\rho\sigma}(\zeta) \Upsilon_{\alpha\beta}(\boldsymbol{\eta}) + \varepsilon a^{\alpha\beta\rho\sigma} \gamma_{\rho\sigma}(\zeta) \gamma_{\alpha\beta}(\boldsymbol{\eta}) \right\} \sqrt{a} dy, \\ L(\boldsymbol{\eta}) &= \int_\omega p^i \eta_i \sqrt{a} dy.\end{aligned}$$

The linear form  $L$  takes into account the applied forces. The given functions  $p^i$  are assumed to be in  $L^2(\omega)$ .

The symmetric bilinear form  $B$  and the linear form  $L$  are continuous over the space  $\mathbf{V}(\omega)$ . Hence the existence and uniqueness of the solution of the above variational problem follow, by the *Lax-Milgram lemma*, from:

**Theorem 1.** Assume that  $\varphi \in \mathcal{C}^3(\bar{\omega})$  and that length  $\gamma_0 > 0$ . There exists a constant  $\beta$  such that

$$\beta > 0 \text{ and } B(\boldsymbol{\eta}, \boldsymbol{\eta}) \geq \beta \|\boldsymbol{\eta}\|_{H^1(\omega) \times H^1(\omega) \times H^2(\omega)}^2$$

for all  $\boldsymbol{\eta} \in \mathbf{V}(\omega)$ . □

Theorem 1 was first established by Bernadou and Ciarlet [3]. The proof relied on various equivalences of norms involving covariant derivatives (also due to Rougée [23]), on a rigid displacement lemma (cf. (iii) below), and on technical inequalities combined with weak lower semi-continuity properties of the associated quadratic functional; an outline of this proof was also given in Ciarlet [6]. A notable simplification of this proof was recently proposed by Ciarlet and Miara [10]. It is this proof that we sketch here; the full, detailed, proof is given in Bernadou *et al* [4].

*Outline of the proof of Theorem 1:*

(i) There exists a constant  $C_1 > 0$  such that

$$a^{\alpha\rho}(y)a^{\beta\sigma}(y)t_{\rho\sigma}t_{\alpha\beta} \geq C_1 \sum_{\alpha,\beta} |t_{\alpha\beta}|^2$$

for all  $y \in \bar{\omega}$  and all symmetric tensors  $(t_{\alpha\beta})$ . Since

$$a^{\alpha\beta}(y)a^{\rho\sigma}(y)t_{\rho\sigma}t_{\alpha\beta} \geq 0$$

on the other hand, it suffices to show that there exists a constant  $C_2 > 0$  such that

$$\left\{ \sum_{\alpha,\beta} \|\Upsilon_{\alpha\beta}(\boldsymbol{\eta})\|_{L^2}^2 + \sum_{\alpha,\beta} \|\gamma_{\alpha\beta}(\boldsymbol{\eta})\|_{L^2}^2 \right\}^{1/2} \geq C_2 \|\boldsymbol{\eta}\|_{H^1 \times H^1 \times H^2},$$

for all  $\boldsymbol{\eta} \in \mathbf{V}(\omega)$ , where, here and subsequently, we let  $L^2 = L^2(\omega)$ ,  $H^m = H^m(\omega)$  at some places, for the sake of conciseness.

(ii) Define the space

$$\mathbf{E}(\omega) = \{\boldsymbol{\eta} = (\eta_i); \eta_\alpha \in L^2, \eta_3 \in H^1, \gamma_{\alpha\beta}(\boldsymbol{\eta}) \in L^2, \Upsilon_{\alpha\beta}(\boldsymbol{\eta}) \in L^2\},$$

where both relations  $\gamma_{\alpha\beta}(\boldsymbol{\eta}) \in L^2$  and  $\Upsilon_{\alpha\beta}(\boldsymbol{\eta}) \in L^2$  are to be understood in the sense of distributions. We show that

$$\mathbf{E}(\omega) = H^1(\omega) \times H^1(\omega) \times H^2(\omega).$$

Let  $\boldsymbol{\eta} = (\eta_i)$  be an arbitrary element of the space  $\mathbf{E}(\omega)$ . The relations

$$e_{\alpha\beta}(\boldsymbol{\eta}) := \frac{1}{2}(\partial_\alpha \eta_\beta + \partial_\beta \eta_\alpha) = \gamma_{\alpha\beta}(\boldsymbol{\eta}) + \Gamma_{\alpha\beta}^\rho \eta_\rho + b_{\alpha\beta} \eta_3$$

imply that the functions  $e_{\alpha\beta}(\boldsymbol{\eta})$  belong to the space  $L^2(\omega)$ . Hence the identities (in the sense of distributions)

$$\partial_{\alpha\beta} \eta_\rho = \partial_\alpha e_{\beta\rho}(\boldsymbol{\eta}) + \partial_\beta e_{\alpha\rho}(\boldsymbol{\eta}) - \partial_\rho e_{\alpha\beta}(\boldsymbol{\eta})$$

show that  $\partial_\rho(\partial_\alpha \eta_\rho) \in H^{-1}(\omega)$  (the assumption  $\Phi \in \mathcal{C}^3(\bar{\omega})$  is used here). Since  $\partial_\alpha \eta_\rho \in H^{-1}(\omega)$  (recall that  $\eta_\rho \in L^2(\omega)$ ), a lemma of J L Lions (first mentioned by Magenes and Stampacchia [19] and proved in Duvaut and Lions ([15], p. 110), then extended to Lipschitz-continuous boundaries in Borchers and Sohr [5] and Amrouche and Girault [2] implies that the distributions  $\partial_\alpha \eta_\rho$  are in the space  $L^2(\omega)$ ; hence  $\eta_\rho \in H^1(\omega)$ . The definition of  $\Upsilon_{\alpha\beta}(\eta_3)$  then implies that  $\eta_3 \in H^2(\omega)$ , and the inclusion

$$\mathbf{E}(\omega) \subset H^1(\omega) \times H^1(\omega) \times H^2(\omega)$$

is thus established; the other inclusion clearly holds.

When equipped with the norm  $\|\cdot\|_{\mathbf{E}}$  defined by

$$\|\boldsymbol{\eta}\|_{\mathbf{E}} = \left\{ \sum_{\alpha,\beta} \|\Upsilon_{\alpha\beta}(\boldsymbol{\eta})\|_{L^2}^2 + \sum_{\alpha,\beta} \|\gamma_{\alpha\beta}(\boldsymbol{\eta})\|_{L^2}^2 + \sum_{\alpha} \|\eta_{\alpha}\|_{L^2}^2 + \|\eta_3\|_{H^1}^2 \right\}^{1/2},$$

the space  $\mathbf{E}(\omega)$  becomes a Hilbert space. Since the identity mapping from the space  $H^1(\omega) \times H^1(\omega) \times H^2(\omega)$  into the space  $\mathbf{E}(\omega)$  is continuous and onto (we just showed that the two spaces are identical), and since both spaces are complete, the open mapping theorem implies that the identity mapping from  $\mathbf{E}(\omega)$  onto  $H^1(\omega) \times H^1(\omega) \times H^2(\omega)$  is also continuous. Hence there exists a constant  $C_3 > 0$  such that the following generalized Korn's inequality holds:

$$\|\boldsymbol{\eta}\|_{\mathbf{E}} \geq C_3 \left\{ \sum_{\alpha} \|\eta_{\alpha}\|_{H^1}^2 + \|\eta_3\|_{H^2}^2 \right\}^{1/2}$$

for all  $\boldsymbol{\eta} \in H^1(\omega) \times H^1(\omega) \times H^2(\omega)$ .

In other words, the norm  $\|\cdot\|_{\mathbf{E}}$  is a norm over the space  $H^1(\omega) \times H^1(\omega) \times H^2(\omega)$ , equivalent to its product norm.

(iii) We next show that the semi-norm  $|\cdot|_{\mathbf{E}}$  defined by

$$|\boldsymbol{\eta}|_{\mathbf{E}} = \left\{ \sum_{\alpha,\beta} \|\Upsilon_{\alpha\beta}(\boldsymbol{\eta})\|_{L^2}^2 + \sum_{\alpha,\beta} \|\gamma_{\alpha\beta}(\boldsymbol{\eta})\|_{L^2}^2 \right\}^{1/2},$$

is a norm over the space  $\mathbf{V}(\omega)$ . To this end, it suffices to show that

$$\boldsymbol{\eta} \in \mathbf{V}(\omega) \text{ and } |\boldsymbol{\eta}|_{\mathbf{E}} = 0 \Rightarrow \boldsymbol{\eta} = 0.$$

The *generalized displacement lemma* (cf. Bernadou and Ciarlet [[3], th. 5.1-1] or Bernadou, *et al* [[4], lemma 2.5]) shows that, if an element  $\boldsymbol{\eta} \in H^1(\omega) \times H^1(\omega) \times H^2(\omega)$  satisfies  $\Upsilon_{\alpha\beta}(\boldsymbol{\eta}) = \gamma_{\alpha\beta}(\boldsymbol{\eta}) = 0$  in  $\omega$ , there exist two vectors  $\mathbf{a} \in \mathbb{R}^3$  and  $\mathbf{b} \in \mathbb{R}^3$  such that

$$\eta_i(y) \mathbf{a}^i(y) = \mathbf{a} + \mathbf{d} \times \Phi(y) \text{ for all } y \in \omega.$$

The conclusion then follows by taking into account the boundary conditions satisfied by the functions  $\eta_i$  along  $\gamma_0$  (the assumption length  $\gamma_0 > 0$  is needed here).

(iv) We finally show that, over the space  $\mathbf{V}(\omega)$ , the norm  $|\cdot|_{\mathbf{E}}$  is in fact equivalent to the product norm  $\|\cdot\|_{H^1 \times H^1 \times H^2}$ , i.e., that there exists a constant  $C_2$  such that the inequality announced in (i) holds. Otherwise, there exists a sequence  $(\boldsymbol{\eta}^k)$  of elements in  $\mathbf{V}(\omega)$  such that

$$|\boldsymbol{\eta}^k|_{\mathbf{E}} \rightarrow 0 \text{ as } k \rightarrow \infty, \|\boldsymbol{\eta}^k\|_{H^1 \times H^1 \times H^2} = 1 \text{ for all } k.$$

By the Rellich–Kondrašov theorem, there exists a subsequence  $(\boldsymbol{\eta}^l)$  that converges in the space  $L^2(\omega) \times L^2(\omega) \times H^1(\omega)$ . Since  $|\boldsymbol{\eta}^l|_E \rightarrow 0$  as  $l \rightarrow \infty$ , the subsequence  $(\boldsymbol{\eta}^l)$  is a Cauchy sequence with respect to  $\|\cdot\|_E$ , whence also to  $\|\cdot\|_{H^1 \times H^1 \times H^2}$  by (ii). Let  $\boldsymbol{\eta} = \lim_{l \rightarrow \infty} \boldsymbol{\eta}^l$  in the space  $V(\omega)$ . On the one hand,  $|\boldsymbol{\eta}|_E = \lim_{l \rightarrow \infty} |\boldsymbol{\eta}^l|_E = 0$ ; on the other,  $\|\boldsymbol{\eta}\|_{H^1 \times H^1 \times H^2} = \lim_{l \rightarrow \infty} \|\boldsymbol{\eta}^l\|_{H^1 \times H^1 \times H^2} = 1$ . Hence we have reached a contradiction, and the proof is complete.  $\square$

*Remarks.* (1) No geometrical assumption on the middle surface  $S$  is needed here (by contrast, an assumption of uniform ellipticity will be introduced to handle the membrane model; cf. § 4). (2) The same analysis can be applied to the two-dimensional shell model of Naghdi [22]; cf. Bernadou *et al* ([3], th. 3.1).  $\square$

### 3. The two-dimensional bending shell model

As observed in Ciarlet [9], Koiter's model is *not* a limit model, i.e., one that can be obtained in a rational fashion as a limit of the three-dimensional equations as  $\varepsilon \rightarrow 0$ . Indeed, Sanchez–Palencia [26] has shown that the solution of the three-dimensional shell equations has two essentially different behaviors as the thickness approaches zero, according to the geometry of the middle surface and to the boundary conditions: It converges either to the solution of the bending shell model, or to the solution of the membrane shell model, which are described in this and the next sections (for more details about this limit behavior, the relations between these models, and the differences between shells and plates, see also Destuynder [14], Sanchez–Palencia [24, 25], Ciarlet [7, 8] Miara and Sanchez–Palencia [20]).

More specifically, let

$$V_0(\omega) = \{\boldsymbol{\eta} \in V(\omega); \gamma_{\alpha\beta}(\boldsymbol{\eta}) = 0 \text{ in } \omega\}$$

denote the space of *inextensional displacements*, where the strain tensor  $(\gamma_{\alpha\beta}(\boldsymbol{\eta}))$  and the space  $V(\omega)$  are defined as in § 1 and § 2, respectively.

If  $V_0(\omega) \neq \{0\}$  (there exist such instances), the first non-zero term  $\zeta = (\zeta_i)$  of a formal asymptotic expansion as powers of  $\varepsilon$  of the covariant components of the three-dimensional displacement is independent of the transverse variable, and it solves the following two-dimensional shell model, called the bending model:

$$\zeta \in V_0(\omega) \text{ and } B_0(\zeta, \boldsymbol{\eta}) = L(\boldsymbol{\eta}) \text{ for all } \boldsymbol{\eta} \in V_0(\omega),$$

where the space  $V_0(\omega)$  is defined as above,

$$B_0(\zeta, \boldsymbol{\eta}) = \int_{\omega} \frac{\varepsilon}{3} a^{\alpha\beta\rho\sigma} \Upsilon_{\rho\sigma}(\zeta) \Upsilon_{\alpha\beta}(\boldsymbol{\eta}) \sqrt{a} \, dy,$$

and the linear form  $L$  has the same expression as in § 2 (the tensors  $(a^{\alpha\beta\rho\sigma})$  and  $(\Upsilon_{\alpha\beta}(\boldsymbol{\eta}))$  are defined as in § 1).

The existence and uniqueness of the solution of the above variational equations are consequences of the following theorem (note that  $V_0(\omega)$  is a closed subspace of  $V(\omega)$ ).

**Theorem 2.** Assume that  $\varphi \in \mathcal{C}^3(\bar{\omega})$ ,  $\text{length } \gamma_0 > 0$ , and  $\forall \gamma \neq \gamma_0$ . Then there exists a constant  $\beta_0$  such that

$$\beta_0 > 0 \text{ and } B_0(\eta, \eta) \geq \beta_0 \|\eta\|_{H^1(\omega) \times H^1(\omega) \times H^2(\omega)}^2$$

for all  $\eta \in V_0(\omega)$ .

*Proof.* By part (i) of the proof of Theorem 1, there exists a constant  $C_4 > 0$  such that

$$B_0(\eta, \eta) \geq C_4 \left\{ \sum_{\alpha, \beta} \|\Upsilon_{\alpha\beta}(\eta)\|_{L^2}^2 \right\}^{1/2} \text{ for all } \eta \in V(\omega).$$

In the same proof, we have seen that the semi-norm  $|\cdot|_E$  is a norm over  $V(\omega)$ , equivalent to the product norm  $\|\cdot\|_{H^1 \times H^1 \times H^2}$  (cf. (iv)). The conclusion thus follows, since

$$\left\{ \sum_{\alpha, \beta} \|\Upsilon_{\alpha\beta}(\eta)\|_{L^2}^2 \right\}^{1/2} = |\eta|_E \text{ for all } \eta \in V_0(\omega). \quad \square$$

#### 4. The two-dimensional membrane shell model

Let the space  $V_0(\omega)$  of inextensional displacements be defined as in § 3. If  $V_0(\omega) = \{0\}$  (there exist such instances), the first non-zero term  $\zeta = (\zeta_i)$  of a formal asymptotic expansion as powers of  $\varepsilon$  of the covariant components of the three-dimensional displacement is independent of the transverse variable, and it solves the following two-dimensional shell model, called the membrane model:

$$\zeta \in V_1(\omega) \text{ and } B_1(\zeta, \eta) = L(\eta) \text{ for all } \eta \in V_1(\omega),$$

where

$$V_1(\omega) = \{\eta = (\eta_i); \eta_\alpha \in H^1(\omega), \eta_3 \in L^2(\omega), \eta_\alpha = 0 \text{ on } \gamma_0\},$$

$$B_1(\zeta, \eta) = \int_{\omega} \varepsilon a^{\alpha\beta\rho\sigma} \gamma_{\rho\sigma}(\zeta) \gamma_{\alpha\beta}(\eta) \sqrt{a} dy,$$

and the linear form  $L$  has the same expression as in § 2 (the tensors  $(a^{\alpha\beta\rho\sigma})$  and  $(\gamma_{\alpha\beta}(\eta))$  are defined as in § 1).

The existence and uniqueness of the solution of the above variational equations are consequences of the following result:

**Theorem 3.** Assume that the boundary  $\gamma$  is of class  $\mathcal{C}^3$  and that  $\gamma_0 = \gamma$ . Assume further that  $\varphi$  is analytic in an open set  $\omega'$  containing  $\bar{\omega}$ . Assume finally that the surface  $S$  is uniformly elliptic, in the sense that there exists a constant  $b$  such that

$$b > 0 \text{ and } b_{\alpha\beta}(y) \xi^\alpha \xi^\beta \geq b |\xi|^2$$

for all  $y \in \bar{\omega}$  and all  $\xi = (\xi^\alpha) \in \mathbb{R}^2$ , where  $(b_{\alpha\beta})$  denotes the second fundamental form of  $S$ . Then there exists a constant  $\beta_1$  such that

$$\beta_1 > 0 \text{ and } B_1(\eta, \eta) \geq \beta_1 \|\eta\|_{H^1(\omega) \times H^1(\omega) \times L^2(\omega)}^2$$

for all  $\eta \in V_1(\omega) = H_0^1(\omega) \times H_0^1(\omega) \times L^2(\omega)$ . □



Two different proofs of Theorem 3 are given in Ciarlet and Sanchez-Palencia [12, 13], and in Ciarlet and Lods [11]. It is the latter proof that we sketch here.

### Outline of the proof of Theorem 3

(i) By part (i) of the proof of Theorem 1, there exists a constant  $C_5 > 0$  such that

$$B_1(\boldsymbol{\eta}, \boldsymbol{\eta}) \geq C_5 \left\{ \sum_{\alpha, \beta} \|\gamma_{\alpha\beta}(\boldsymbol{\eta})\|_{L^2}^2 \right\}^{1/2}$$

for all  $\boldsymbol{\eta} \in H^1 \times H^1 \times L^2$ . It thus suffices to show that there exists a constant  $C_6 > 0$  such that

$$\left\{ \sum_{\alpha, \beta} \|\gamma_{\alpha\beta}(\boldsymbol{\eta})\|_{L^2}^2 \right\}^{1/2} \geq C_6 \|\boldsymbol{\eta}\|_{H^1 \times H^1 \times L^2}$$

for all  $\boldsymbol{\eta} = (\eta_i) \in H_0^1 \times H_0^1 \times L^2$  (recall that we assume  $\gamma_0 = \gamma$ ).

(ii) Using the same arguments as in part (ii) of the proof of Theorem 1, i.e., in particular the lemma of J L Lions and the open mapping theorem, one successively shows that

$$\{\boldsymbol{\eta} = (\eta_i); \eta_i \in L^2, \gamma_{\alpha\beta}(\boldsymbol{\eta}) \in L^2\} = H^1 \times H^1 \times L^2,$$

and that there exists a constant  $C_7 > 0$  such that the following *generalized Korn's inequality* holds:

$$\left\{ \sum_{\alpha, \beta} \|\gamma_{\alpha\beta}(\boldsymbol{\eta})\|_{L^2}^2 + \sum_i \|\eta_i\|_{L^2}^2 \right\}^{1/2} \geq C_7 \left\{ \sum_{\alpha} \|\eta_{\alpha}\|_{H^1}^2 + \|\eta_3\|_{L^2}^2 \right\}^{1/2}$$

for all  $\boldsymbol{\eta} \in H^1 \times H^1 \times L^2$ .

(iii) One next establishes another rigid displacement lemma: If the surface  $S$  is uniformly elliptic (this assumption is needed from now on), the space

$$\mathbf{R}(\omega) = \{\boldsymbol{\eta} = (\eta_i) \in H_0^1 \times H_0^1 \times L^2; \gamma_{\alpha\beta}(\boldsymbol{\eta}) = 0 \text{ in } \omega\}$$

is *finite-dimensional*. To this end, one first observes that, if  $\zeta = (\zeta_i) \in \mathbf{R}(\omega)$  then  $\tilde{\zeta} := (\zeta_{\alpha}) \in H_0^1 \times H_0^1$  solves the variational equations

$$A_0(\zeta, \boldsymbol{\eta}) + A_1(\tilde{\zeta}, \tilde{\boldsymbol{\eta}}) = 0 \text{ for all } \tilde{\boldsymbol{\eta}} := (\eta_{\alpha}) \in H_0^1 \times H_0^1,$$

where

$$\begin{aligned} A_0(\zeta, \boldsymbol{\eta}) &= \int_{\omega} \left\{ \frac{b_{22}}{b_{11}} \partial_1 \zeta_1 \partial_1 \eta_1 + \left( \partial_2 \zeta_1 - 2 \frac{b_{12}}{b_{11}} \partial_1 \zeta_1 \right) \partial_2 \eta_1 \right\} dy \\ &\quad + \int_{\omega} \left\{ \frac{b_{11}}{b_{22}} \partial_2 \zeta_2 \partial_2 \eta_2 + \left( \partial_1 \zeta_2 - 2 \frac{b_{12}}{b_{22}} \partial_2 \zeta_2 \right) \partial_1 \eta_2 \right\} dy, \\ A_1(\tilde{\zeta}, \tilde{\boldsymbol{\eta}}) &= \int_{\omega} \left\{ \left( \Gamma_{22}^{\rho} - \frac{b_{22}}{b_{11}} \Gamma_{11}^{\rho} \right) \zeta_{\rho} \partial_1 \eta_1 + 2 \left( \frac{b_{12}}{b_{11}} \Gamma_{11}^{\rho} - \Gamma_{12}^{\rho} \right) \zeta_{\rho} \partial_2 \eta_1 \right\} dy \\ &\quad + \int_{\omega} \left\{ \left( \Gamma_{11}^{\rho} - \frac{b_{11}}{b_{22}} \Gamma_{22}^{\rho} \right) \zeta_{\rho} \partial_2 \eta_2 + 2 \left( \frac{b_{12}}{b_{22}} \Gamma_{22}^{\rho} - \Gamma_{12}^{\rho} \right) \zeta_{\rho} \partial_1 \eta_2 \right\} dy. \end{aligned}$$

One next shows that there exists a constant  $C_9 > 0$  such that

$$A_0(\tilde{\eta}, \tilde{\eta}) \geq C_8 \|\tilde{\eta}\|_{H^1 \times H^1}^2, \text{ for all } \tilde{\eta} = (\eta_\alpha) \in H_0^1 \times H_0^1.$$

Since there exists a constant  $C_9 > 0$  such that

$$A_1(\tilde{\eta}, \tilde{\eta}) \geq -2C_8 C_9 \|\tilde{\eta}\|_{L^2 \times L^2} \|\tilde{\eta}\|_{H^1 \times H^1}, \text{ for all } \tilde{\eta} \in H^1 \times H^1,$$

one easily deduces that the operator  $T$  from  $L^2 \times L^2$  into  $H_0^1 \times H_0^1$  defined by the relations

$$A_0(Tq, \eta) + A_1(Tq, \eta) + \lambda \int_{\omega} (Tq)^\alpha \eta_\alpha dy = \int_{\omega} q^\alpha \eta_\alpha dy, \text{ where } \lambda = 2C_8(C_9)^2,$$

for all  $\eta = (\eta_\alpha) \in H_0^1 \times H_0^1$ , is compact.

Since  $\tilde{\zeta} \in \text{Ker}(I - \lambda T)$  and  $\text{Ker}(I - \lambda T)$  is finite-dimensional ( $T$  is compact), the assertion is proved.

(iv) By refining the regularity assumptions on the boundary  $\gamma$  (which was so far assumed to be only Lipschitz-continuous) and on the mapping  $\phi$  (which was so far assumed to be of class  $\mathcal{C}^2$  on  $\bar{\omega}$ ), we can strengthen the result of part (iii). More specifically, we show that, if  $\gamma$  is of class  $\mathcal{C}^3$ , and  $\phi$  is analytic in an open set  $\omega'$  containing  $\bar{\omega}$ , the space  $\mathbf{R}(\omega)$  (as defined in (iii)) reduces to  $\{0\}$ .

Consider the boundary-value problem:

$$[\gamma_{11}(\zeta) := ]\partial_1 \zeta_1 - \Gamma_{11}^\rho \zeta_\rho - b_{11} \zeta_3 = 0 \text{ in } \omega,$$

$$[\gamma_{12}(\zeta) := ]\frac{1}{2}\partial_2 \zeta_1 + \frac{1}{2}\partial_1 \zeta_2 - \Gamma_{12}^\rho \zeta_\rho - b_{12} \zeta_3 = 0 \text{ in } \omega,$$

$$[\gamma_{22}(\zeta) := ]\partial_2 \zeta_2 - \Gamma_{22}^\rho \zeta_\rho - b_{22} \zeta_3 = 0 \text{ in } \omega, \zeta_1 = 0 \text{ on } \gamma.$$

This first-order system is a uniformly elliptic system (the assumption of uniform ellipticity of  $S$  is needed here) that satisfies the supplementary condition on  $L$ , and  $\zeta_1 = 0$  on  $\gamma$  is a complementing boundary condition, in the sense of Agmon *et al* [1] (this was first observed by Geymonat and Sanchez-Palencia [16]).

We thus need to show that  $\zeta = 0$  is the only solution to the boundary value problem:

$$\begin{cases} \gamma_{\alpha\beta}(\zeta) = 0 \text{ in } \omega, \\ \zeta_\alpha = 0 \text{ on } \gamma. \end{cases}$$

Since the boundary  $\gamma$  is not a characteristic curve for the reduced Cauchy problem where  $\zeta_3$  has been eliminated, Holmgren's uniqueness theorem shows that  $\zeta = 0$  is the only solution in a small enough neighborhood of any point of  $\gamma$  (the coefficients are analytic in  $\omega'$  because  $\phi$  is analytic in  $\omega'$ ).

By a result of Morrey and Nirenberg [21], any solution of a uniformly elliptic system whose coefficients are analytic in  $\omega$  is analytic in  $\omega$ . Therefore  $\zeta = 0$  is the only solution in  $\omega$ , by the analytic continuation theorem for analytic functions of several real variables.

(v) In order to conclude, it suffices to show that there exists a constant  $C_{10} > 0$  such

that

$$\left\{ \sum_{\alpha, \beta} \|\gamma_{\alpha\beta}(\boldsymbol{\eta})\|_{L^2}^2 \right\}^{1/2} \geq C_{10} \left\{ \sum_i \|\eta_i\|_{L^2}^2 \right\}^{1/2} \text{ for all } \boldsymbol{\eta} \in H_0^1 \times H_0^1 \times L^2,$$

as the desired inequality (that involving the constant  $C_6$ , cf. part (i)) will then follow from the inequality established in part (ii).

If this assertion is false, there exists a sequence  $(\boldsymbol{\eta}^k)$  of elements in  $H_0^1 \times H_0^1 \times L^2$  such that

$$\left\{ \sum_{\alpha, \beta} \|\gamma_{\alpha\beta}(\boldsymbol{\eta}^k)\|_{L^2}^2 \right\}^{1/2} \rightarrow 0 \text{ as } k \rightarrow \infty \text{ and } \left\{ \sum_i \|\eta_i^k\|_{L^2}^2 \right\}^{1/2} = 1 \text{ for all } k.$$

Hence there exists a subsequence  $(\boldsymbol{\eta}^l)$  and an element  $\boldsymbol{\eta} = (\eta_i) \in H_0^1 \times H_0^1 \times L^2$  such that ( $\rightarrow$  and  $\rightharpoonup$  denote strong and weak convergences, respectively):

$$\eta_\alpha^l \rightharpoonup \eta_\alpha \text{ in } H_0^1, \eta_\alpha^l \rightarrow \eta_\alpha \text{ in } L^2, \eta_3^l \rightarrow \eta_3 \text{ in } L^2.$$

Since  $\gamma_{\alpha\beta}(\boldsymbol{\eta}^l) \rightarrow \gamma_{\alpha\beta}(\boldsymbol{\eta})$  in  $L^2$  on the one hand and  $\gamma_{\alpha\beta}(\boldsymbol{\eta}^l) \rightarrow 0$  in  $L^2$  on the other, we first conclude that  $\boldsymbol{\eta} = \mathbf{0}$ , by (iv).

The convergences  $\gamma_{\alpha\beta}(\boldsymbol{\eta}^l) \rightarrow 0$  in  $L^2$  and  $\eta_\alpha^l \rightarrow 0$  in  $L^2$ , combined with the definition of the functions  $\gamma_{\alpha\beta}(\boldsymbol{\eta})$  imply that  $(b_{11} \in C^0(\bar{\omega}))$  does not vanish in  $\bar{\omega}$ , by the assumed uniform ellipticity of  $S$ )

$$\partial_2 \eta_1^l + \partial_1 \eta_2^l - 2 \frac{b_{12}}{b_{11}} \partial_1 \eta_1^l \rightarrow 0 \text{ in } L^2,$$

$$\partial_2 \eta_2^l - \frac{b_{22}}{b_{11}} \partial_1 \eta_1^l \rightarrow 0 \text{ in } L^2.$$

From these convergences and the relations

$$\int_{\omega} \partial_2 \eta_1^l \partial_1 \eta_2^l dy = \int_{\omega} \partial_1 \eta_1^l \partial_2 \eta_2^l dy,$$

we then infer that

$$\int_{\omega} \left\{ \left( \partial_2 \eta_1^l - \frac{b_{12}}{b_{11}} \partial_1 \eta_1^l \right)^2 + \frac{1}{(b_{11})^2} (b_{11} b_{22} - (b_{12})^2) (\partial_1 \eta_1^l)^2 \right\} dy \rightarrow 0.$$

This last convergence, combined with the uniform ellipticity of  $S$ , then implies that

$$\eta_3^l = \frac{1}{b_{11}} \partial_1 \eta_1^l - \frac{1}{b_{11}} (\partial_1 \eta_1^l - b_{11} \eta_3^l) \rightarrow 0 \text{ in } L^2.$$

Hence  $\boldsymbol{\eta}^l \rightarrow \mathbf{0}$  in  $L^2 \times L^2 \times L^2$ , which contradicts  $\left\{ \sum_i \|\eta_i^l\|_{L^2}^2 \right\}^{1/2} = 1$ , and the proof of Theorem 3 is complete.  $\square$

## References

- [1] Agmon S, Douglis A and Nirenberg L, Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions II, *Commun. Pure Appl. Math.* **17** (1964) 35–92
- [2] Amrouche C and Girault V, Propriétés fonctionnelles d'opérateurs; applications au problème de Stokes en dimension quelconque, Report R90025, Laboratoire d'Analyse Numérique, Université Pierre et Marie Curie, Paris (1990)
- [3] Bernadou M and Ciarlet P G, Sur l'ellipticité du modèle linéaire de coques de W T Koiter, in *Computing Methods in Applied Sciences and Engineering* (eds) R Glowinski and J L Lions, Lecture Notes in Economics and Mathematical Systems, (Heidelberg: Springer-Verlag) Vol. 134 (1976) pp. 89–136
- [4] Bernadou M, Ciarlet P G and Miara B, Existence theorems for two-dimensional linear shell theories, *J. Elast.* (1993) to appear
- [5] Borchers W and Sohr H, On the equations  $\operatorname{rot} v = g$  and  $\operatorname{div} u = f$  with zero boundary conditions, *Hokkaido Math. J.* **19** (1990) 67–87
- [6] Ciarlet P G, On questions of existence in shell theory, *J. Indian Math. Soc.* **40** (1976) 131–143
- [7] Ciarlet P G, *Plates and Junctions in elastic multi-structures: An asymptotic analysis*, Masson, Paris, (1990) (Heidelberg: Springer-Verlag)
- [8] Ciarlet P G, Exchange de limites en théorie asymptotique de coques. I. En premier lieu, "la coque devient une plaque", *C. R. Acad. Sci. Paris, Sér. I*, **315** (1992a) 107–111
- [9] Ciarlet P G, Exchange de limites en théorie asymptotique de coques. II. En premier lieu, l'épaisseur tend vers zéro, *C. R. Acad. Sci. Paris, Sér. I*, **315** (1992) 227–233
- [10] Ciarlet P G and Miara B, On the ellipticity of linear shell models, *Z. Angew. Math. Phys.* **43** (1992) 243–253
- [11] Ciarlet P G and Lods V, On the ellipticity of linear membrane shell equations (1994) to appear
- [12] Ciarlet P G and Sanchez-Palencia E, Un théorème d'existence et d'unicité pour les équations des coques membranaires, *C. R. Acad. Sci. Paris, Sér. I*, **317** (1993), 801–805
- [13] Ciarlet P G and Sanchez-Palencia E, An existence and uniqueness theorem for the two-dimensional linear membrane shell equations, (1994) to appear
- [14] Destuynder P, A classification of thin shell theories, *Acta Applicandae Mathematicae* **4** (1985) 15–63
- [15] Duvaut G and Lions J L, *Les Inéquations en Mécanique et en Physique* (Paris: Dunod)
- [16] Geymonat G and Sanchez-Palencia E, Remarques sur la rigidité infinitésimale de certaines surfaces elliptiques non régulières, non convexes et applications, *C. R. Acad. Sci. Paris, Sér. I*, **313** (1991) 645–651
- [17] John F, Estimates for the derivatives of the stresses in a thin shell and interior shell equations, *Commun. Pure Appl. Math.* **18** (1965) 235–267
- [18] Koiter W T, On the foundations of the linear theory of thin elastic shells, *Proc. Kon. Ned. Akad. Wetensch.* **B73** (1970) 169–195
- [19] Magenes E and Stampacchia G, I problemi al contorno per le equazioni differenziali di tipo ellittico, *Ann. Scuola Norm. Sup. Pisa* **12** (1958) 247–358
- [20] Miara B and Sanchez-Palencia E (1994), To appear
- [21] Morrey C B and Nirenberg L, On the analyticity of the solution of linear elliptic systems of partial differential equations, *Commun. Pure Appl. Math.* **10** (1957) 271–290
- [22] Naghdi P M, Foundations of elastic shell theory, in *Progress in Solid Mechanics* (Amsterdam: North-Holland) Vol. 4 (1963) pp. 1–90
- [23] Rougée P, *Equilibre des Coques Elastiques Minces Inhomogènes en Theorie Non Linéaire*, Doctoral Dissertation, Université de Paris (1969)
- [24] Sanchez-Palencia E, Statique et dynamique des coques minces. I. Cas de flexion pure non inhibée, *C. R. Acad. Sci. Paris, Sér. I*, **309** (1989) 411–417
- [25] Sanchez-Palencia E, Statique et dynamique des coques minces. II. Cas de flexion pure inhibée, *C. R. Acad. Sci. Paris, Sér. I*, **309** (1989) 531–537
- [26] Sanchez-Palencia E, Passage à la limite de l'élasticité tridimensionnelle à la théorie asymptotique des coques minces, *C. R. Acad. Sci. Paris, Sér. II*, **311** (1990) 909–916

# Absolutely expedient algorithms for learning Nash equilibria

V V PHANSALKAR, P S SASTRY and M A L THATHACHAR

Department of Electrical Engineering, Indian Institute of Science, Bangalore 560 012, India

Dedicated to the memory of Professor K G Ramanathan

**Abstract.** This paper considers a multi-person discrete game with random payoffs. The distribution of the random payoff is unknown to the players and further none of the players know the strategies or the actual moves of other players. A class of absolutely expedient learning algorithms for the game based on a decentralised team of Learning Automata is presented. These algorithms correspond, in some sense, to rational behaviour on the part of the players. All stable stationary points of the algorithm are shown to be Nash equilibria for the game. It is also shown that under some additional constraints on the game, the team will always converge to a Nash equilibrium.

**Keywords.** Nash equilibria; Decentralised learning algorithm.

## 1. Introduction

This paper is concerned with a learning problem in a general multiperson stochastic game with incomplete information. We study a class of decentralised algorithms for learning Nash equilibria. For this purpose, we employ team concepts associated with Learning Automata models [13].

The game we consider is a discrete stochastic game played by  $N$  players. Each of the players has finitely many actions one of which he plays at each instant. After each play, the payoffs to individual players are random variables. The objective for each player is to maximise his expected payoff. Further, the game is one of incomplete information [6]. Thus, nothing is known regarding the distributions of the random payoffs. For learning optimal strategies, the game is played repeatedly. We are interested in (asymptotically) learning equilibrium strategies, in the sense of Nash, with respect to the expected value of the payoff. Our interest will be in decentralised learning algorithms. Hence, after each play, each of the players updates his strategy based solely on his current action or move and his payoff. None of the players has any information regarding the other players. As a matter of fact, none of the players need to even know the existence of other players. Thus the game we tackle is also of imperfect information [6].

Such games are useful in tackling problems in many areas such as decentralised control, optimisation, pattern recognition and computer vision. Some of the applications of the game model considered in this paper are discussed in [14]. In many such problems Nash equilibria, in fact, represent the desired solutions. (For a good discussion on the rationality of Nash equilibria see [4, Ch. 2]).

We use a team of learning automata [13] for evolving to the optimal strategies. Games of learning automata have been used as models for adaptive decision making

in many applications [17, 15, 18]. In Learning Automata theory, algorithms for learning optimal strategies have been developed for many special types of finite stochastic games. Some of the models considered are: Two-person zero-sum games [9],  $N$ -person games with common payoff [17, 16, 20] and non-cooperative games such as Prisoner's Dilemma and Stackelberg games [19]. In [14], it is shown that a team of Learning Automata involved in a general  $N$ -person stochastic game will converge to a Nash Equilibrium if each of the team members makes use of a linear algorithm called the  $L_{R-J}$  algorithm [13]. This requires that every member of the team has to use the same algorithm (though may be with different learning parameters). While this may be useful for applications such as optimization, for general  $N$ -person games (that include non-cooperative games) this is a restrictive condition. Here, we expand the earlier result [14] to a case where different players use different algorithms (not necessarily linear) though we require that every player satisfy the 'absolute expediency' property [13]. Informally speaking, the learning algorithm used by a player is absolutely expedient if the algorithm ensures that the expected value of his payoff will increase monotonically with time (when all the other players play to a fixed strategy). We feel that this is a reasonable restriction because, assuming rational behaviour, each player should try to increase his payoff.

We begin by formulating the learning problem in §2. Section 3 gives a brief introduction to the necessary ideas from Learning Automata theory. Section 4 presents the learning algorithm and its analysis. Section 5 concludes the paper with a discussion of the results presented in the paper.

## 2. Problem formulation

In this section we introduce our notation and derive a few results regarding Nash equilibria which will be used later on in the analysis of our algorithm. Most of the formulation in this section can be found in any standard book on Game Theory (e.g., [5, 2, 7]).

Consider a  $N$ -person stochastic game. Each player  $i$  has a finite set of actions or pure strategies,  $S_i$ ,  $1 \leq i \leq N$ . Let cardinality of  $S_i$  be  $m_i$ ,  $1 \leq i \leq N$ . (It may be noted that the sets  $S_i$ ,  $1 \leq i \leq N$ , need not have any common elements and we assume no structure on these sets). Each play of the game consists of each of the players choosing an action. The result of each play is a random payoff to each player. Let  $r_i$  denote the random payoff to player  $i$ ,  $1 \leq i \leq N$ . It is assumed that  $r_i \in [0, 1]$ . Define functions  $d^n: \Pi_{j=1}^N S_j \rightarrow [0, 1]$ ,  $1 \leq i \leq N$ , by

$$d^n(a_1, \dots, a_N) = E[r_i | \text{player } j \text{ chose action } a_j, a_j \in S_j, 1 \leq j \leq N] \quad (1)$$

The function  $d^i$  is called the payoff function or utility function of player  $i$ ,  $1 \leq i \leq N$ . The objective of each player is to maximise his payoff. A strategy for player  $i$  is defined to be a probability vector  $q_i = [q_{i1}, \dots, q_{im}]^t$ , where player  $i$  chooses action  $j$  with probability  $q_{ij}$ . The strategy of a player can be time varying as it would be, for example, during learning. Each of the pure strategies or actions of the  $i$ th player can be considered as a strategy. Let  $e_i$  be a unit probability vector (of appropriate dimension) with  $i$ th component unity and all others zero. Then  $e_i$  is the strategy corresponding to the action  $i$ . (It may be noted that any unit probability vector

represents a pure strategy). A strategy that does not necessarily correspond to a pure strategy is called a mixed strategy.

Given the actions chosen by the players, (1) specifies the expected payoff. We can easily extend the functions  $d^i$ , to the set of all strategies. If  $g^i$  is this extension, then it is defined as

$$\begin{aligned} g^i(\mathbf{q}_1, \dots, \mathbf{q}_N) &= E[r_i | j\text{th player employs strategy } \mathbf{q}_j, 1 \leq j \leq N] \\ &= \sum_{j_1, \dots, j_N} d^i(j_1, \dots, j_N) \prod_{s=1}^N q_{sj_s} \end{aligned} \quad (2)$$

### DEFINITION 2.1

The  $N$ -tuple of strategies  $(\mathbf{q}_1^0, \dots, \mathbf{q}_N^0)$  is said to be a Nash equilibrium, if for each  $i$ ,  $1 \leq i \leq N$ , we have

$$g^i(\mathbf{q}_1^0, \dots, \mathbf{q}_{i-1}^0, \mathbf{q}_i^0, \mathbf{q}_{i+1}^0, \dots, \mathbf{q}_N^0) \geq g^i(\mathbf{q}_1^0, \dots, \mathbf{q}_{i-1}^0, \mathbf{q}, \mathbf{q}_{i+1}^0, \dots, \mathbf{q}_N^0) \quad \forall \text{ probability vectors } \mathbf{q} \in [0, 1]^{m_i} \quad (3)$$

In general, each  $\mathbf{q}_i^0$  above will be a mixed strategy and then we refer to  $(\mathbf{q}_1^0, \dots, \mathbf{q}_N^0)$ , satisfying (3), as a Nash equilibrium in mixed strategies. Every  $N$ -person game will have at least one Nash equilibrium in mixed strategies [4, 5].

We say we have a Nash equilibrium in pure strategies if  $(\mathbf{q}_1^0, \dots, \mathbf{q}_N^0)$  is a Nash equilibrium with each  $\mathbf{q}_i^0$  a unit probability vector. In view of (2), for verifying a Nash equilibrium in pure strategies, we can simplify the condition (3) as given in the definition below.

### DEFINITION 2.2

The  $N$ -tuple of actions  $(a_1^0, \dots, a_N^0)$  (or equivalently the set of strategies  $(\mathbf{e}_{a_1^0}, \dots, \mathbf{e}_{a_N^0})$ ) is called a **Nash equilibrium in pure strategies** if for each  $i$ ,  $1 \leq i \leq N$ ,

$$d^i(a_1^0, \dots, a_{i-1}^0, a_i^0, a_{i+1}^0, \dots, a_N^0) \geq d^i(a_1^0, \dots, a_{i-1}^0, a_i, a_{i+1}^0, \dots, a_N^0), \quad \forall a_i \in S_i. \quad (4)$$

Here, for all  $j$ ,  $a_j^0 \in S_j$ , the set of pure strategies of player  $j$ .

### DEFINITION 2.3

A Nash equilibrium is said to be **strict** if the inequalities in (3) (equivalently, (4), for pure strategies) are strict.

Since each of the sets  $S_i$  is finite, each of the functions,  $d^i: \Pi_{j=1}^N S_j \rightarrow [0, 1]$ , can be represented by a hyper matrix of dimension  $m_1 \times \dots \times m_N$ . These  $N$  hyper matrices together constitute what can be called the reward structure of the game. Since the game is one of incomplete information these payoff matrices are unknown. Now the learning problem for the game can be stated as follows.

Let  $G$  be a  $N$ -person stochastic game with incomplete information. At any instant  $k$ , let the strategy employed by the  $i$ th player be  $\mathbf{q}_i(k)$ . Let  $a_i(k)$  and  $r_i(k)$  be the actual actions selected by  $i$  and the pay off received by  $i$  respectively at  $k$ ,  $k = 0, 1, 2, \dots$ . Find a decentralised learning algorithm for the players (that is, design functions  $T_i$ , where  $\mathbf{q}_i(k+1) = T_i(\mathbf{q}_i(k), a_i(k), r_i(k))$ ) such that  $\mathbf{q}_i(k) \rightarrow \mathbf{q}_i^0$  as  $k \rightarrow \infty$  where  $(\mathbf{q}_1^0, \dots, \mathbf{q}_N^0)$  is a

In § 4, we present a team of Learning Automata model for solving this problem after briefly introducing the concept of a Learning Automaton in § 3. In the remaining part of this section, we state a simple result regarding Nash equilibria, which is needed for the analysis later on. Define  $\mathbf{K} \subset [0, 1]^{m_1 + \dots + m_N}$  by

$$\mathbf{K} = \{Q \in [0, 1]^{m_1 + \dots + m_N} : Q = (q_1, \dots, q_N), \text{ and } \forall i, 1 \leq i \leq N, \\ q_i \text{ is a } m_i\text{-dimensional probability vector}\} \quad (5)$$

It is easy to see that  $\mathbf{K}$  is the set of all  $N$ -tuples of mixed strategies or the set of possible strategies for the game. Let  $\mathbf{K}^* \subset \mathbf{K}$  denote the set of possible pure strategies for the game.  $\mathbf{K}^*$  is defined by

$$\mathbf{K}^* = \{Q \in [0, 1]^{m_1 + \dots + m_N} : Q = (q_1, \dots, q_N), \text{ and } \forall i, 1 \leq i \leq N, q_i \text{ is a} \\ m_i - \text{dimensional probability vector with one component unity}\} \quad (6)$$

It is easy to see that  $\mathbf{K}^*$  can be put in one to one correspondence with the set  $\prod_{j=1}^N S_j$ . Hence we can think of the function  $d^i$ , given by (1) as defined on  $\mathbf{K}^*$ . Similarly, functions  $g^i$ , given by (2), are defined over  $\mathbf{K}$ . Define functions  $h_{is}$ ,  $1 \leq s \leq m_i$ ,  $1 \leq i \leq N$ , on  $\mathbf{K}$  by

$$h_{is}(Q) = E[r_i | \text{player } j \text{ employed strategy } q_j, 1 \leq j \leq N, j \neq i, \\ \text{and player } i \text{ chose action } s] \\ = \sum_{j_1, \dots, j_{i-1}, j_{i+1}, \dots, j_N} d(j_1, \dots, j_{i-1}, s, j_{i+1}, \dots, j_N) \prod_{t \neq s} q_{tj_t} \quad (7)$$

where  $Q = (q_1, \dots, q_N)$ . From (2) and (7), we have

$$\sum_{s=1}^{m_i} h_{is}(Q) q_{is} = g^i(Q) \quad (8)$$

*Lemma 2.1. Any  $Q^0 = (q_1^0, \dots, q_N^0) \in \mathbf{K}$  is a Nash equilibrium if and only if*

$$h_{is}(Q^0) \leq g^i(Q^0), \quad \forall s, i.$$

### COROLLARY 2.1

*Let  $Q^0 = (q_1^0, \dots, q_N^0)$  be a Nash equilibrium. Then for each  $i$ ,*

$$h_{is}(Q^0) = g^i(Q^0) \forall s \text{ such that } q_{is}^0 > 0.$$

Both the above results follow directly from the definition of Nash equilibria and are standard results in Game theory (see, for example, [7, Thm. 3.1], and [2, Ch. 3]).

## 3. Learning Automata

In this section, a brief introduction to Learning Automata [13] models is given.

The basic Learning Automata system consists of a Learning Automation interacting



with an environment. At each instant, the Learning Automaton chooses from a finite set of possible actions and outputs it to the environment. The environment then responds with a signal which evaluates the action. This signal is known as the **Scalar Reinforcement Signal** (SRS). This is a real number and a higher value of the SRS indicates better performance of the system. The SRS is assumed to be stochastic in nature. Otherwise each action can be tried once and the action which returns the highest value of the SRS be selected as the optimal action.

The environment is defined by the tuple  $\langle A, R, \mathcal{F} \rangle$  where

1.  $A$  is the set of actions. It is assumed to be a finite set and

$$A = \{a_1, \dots, a_m\}.$$

2.  $R$  is the set of values the SRS can take. In our case  $R$  will be a subset of the closed interval  $[0, 1]$ . We denote the value of the SRS at instant  $k$  by  $r(k)$ .
3.  $\mathcal{F} = \{F_1, \dots, F_m\}$  is a set of probability distributions over  $R$ .  $F_i$  is the distribution of the SRS, given that the action taken by the system is  $a_i$ .

When the  $F_i$ 's are independent of time, the environment is said to be a stationary environment. We consider such environments in this section. Define

$$d_i = E^i(r)$$

where  $E^i$  denotes expectation with respect to  $F_i$ . Thus,  $d_i$  is the expected value of the SRS if action  $a_i$  is the output. The optimal action is defined as the action which has the highest expected value of the SRS. It is seen that the optimal action is defined independently of the system used to learn it. Thus, the problem is completely defined by the environment.

The basic model of the learning automaton maintains a probability distribution over the set of actions  $A$ . The notation used is the same as the used in describing the environment above. The number of actions is  $m$  and the probability distribution is a  $m$ -dimensional real vector  $p$  such that

$$p = (p_1, \dots, p_m)^t$$

$$p_i \geq 0 \quad \forall i \quad 1 \leq i \leq m$$

$$\sum_{i=1}^m p_i = 1$$

At instant  $k$ , if  $p(k)$  is the action probability vector, the probability of choosing the  $i$ th action is  $p_i(k)$ . Formally, a learning automaton is defined by the tuple  $\langle A, Q, R, T \rangle$  where

1.  $A$  is the (finite) set of actions available to the automaton.
2.  $Q$  is the state of the learning automaton. In standard learning automata theory,  $Q$  is the action probability vector  $p$ . In estimator algorithms [17],  $Q$  consists of the action probability vector along with a vector of estimates. In the algorithms considered in this paper  $Q = p$ . Hence we use  $p$  for the state of the learning automaton.
3.  $R$  is the set from which the SRS takes its values. It is the same as  $R$  in the definition of the environment.

4.  $T$  is the learning algorithm which updates  $p$ .

$$p(k+1) = T(p(k), r(k), a(k))$$

$a(k)$  is the action of the automation at instant  $k$ .

The optimal action is defined to be that which has the highest value of  $d_i$ . One method of evaluating the performance of the system is to check whether the probability of choosing the optimal action becomes unity asymptotically. This is difficult to ascertain and various other performance measures have been defined [13]. They are briefly described below.

Any algorithm should at least do better than a method which chooses actions randomly (with equal probability) at each instant. An automaton which uses this technique is called the pure chance automaton. For comparing the performance of automata with the pure chance automaton, the average value of the SRS at a state  $p$  is needed. This is denoted by  $M(k)$ . Thus

$$M(k) = E[r(k)|p(k)]$$

For the pure chance automaton, this quantity is a constant, denoted by  $M_0$ .

$$M_0 = \frac{1}{m} \sum_{i=1}^m d_i$$

An automaton should at least do better than  $M_0$ . This property is known as expediency.

### DEFINITION 3.1

*A learning automaton is said to be expedient if*

$$\liminf_{k \rightarrow \infty} E[M(k)] > M_0$$

This property does not say much. What is really needed is that the automaton converge to the optimal action. Let the actions of the automaton be  $\{a_1, \dots, a_m\}$ . Let  $a_s$  be the unique optimal action. That is,  $d_s > d_i$  for all  $i \neq s$ .

### DEFINITION 3.2

*A learning automaton is said to be optimal if*

$$\lim_{k \rightarrow \infty} E[M(k)] = d_s$$

Optimality is not an easy property to prove. In fact, no algorithm has this property. A weaker property is that of  $\epsilon$ -optimality. This says that even though the optimal value  $d_s$  cannot be achieved, it should be possible to approach it within any prespecified accuracy.

## DEFINITION 3.3

A learning automaton is said to be  $\varepsilon$ -optimal if

$$\liminf_{k \rightarrow \infty} E[M(k)] > d_s - \varepsilon$$

is achieved for any  $\varepsilon > 0$  by an appropriate choice of the automaton parameters.

In general, different values of the automaton parameters would be required for different values of  $\varepsilon$ .

A property which can be checked without asymptotic analysis and has been shown to imply  $\varepsilon$ -optimality in stationary environments [13] is absolute expediency.

## DEFINITION 3.4

A learning automaton is said to be absolutely expedient if

$$E[M(k+1)|p(k)] > M(k)$$

for all  $k$ , all probabilities in the open interval  $(0, 1)$  and all stationary environments with a unique optimal action.

Necessary and sufficient conditions for an algorithm to be absolutely expedient were first given in [10] and generalised in [1].

Various algorithms have been developed for learning automata. The class of algorithms considered in this paper is those of absolutely expedient algorithms. These algorithms are described below.

$p(k)$  is the action probability vector at instant  $k$ . Let  $a(k)$  denote the action at instant  $k$  and  $r(k)$  the SRS. It is necessary that the SRS take values from a subset of the closed interval  $[0, 1]$ . The general form of the absolutely expedient algorithm is

If the action chosen at instant  $k$  is  $a(k) = a_j$  then

$$\begin{aligned} p_s(k+1) &= p_s(k) - br(k)\alpha_{js}(p(k)) + b(1-r(k))\beta_{js}(p(k)) \quad s \neq j \\ p_j(k+1) &= p_j(k) + br(k) \sum_{s \neq j} \alpha_{js}(p(k)) - b(1-r(k)) \sum_{s \neq j} \beta_{js}(p(k)) \end{aligned} \quad (9)$$

where  $0 < b < 1$  is the learning parameter. The following conditions are imposed on the  $\alpha$  and  $\beta$  functions so that  $p(k+1)$  remains a probability vector.

$$\begin{aligned} \alpha_{js}(p) &\leq p_s \quad \forall j, s \\ \sum_{s \neq j} \beta_{js}(p) &\leq p_j \quad \forall j \end{aligned} \quad (10)$$

Necessary and sufficient conditions for this algorithm to be absolutely expedient were given by Aso-Kimura [1]. These are

$$\begin{aligned} \sum_{s \neq j} p_s \alpha_{sj} &= \sum_{s \neq j} p_j \alpha_{js} \quad \forall j \\ \sum_{s \neq j} p_s \beta_{sj} &= \sum_{s \neq j} p_j \beta_{js} \quad \forall j \end{aligned} \quad (11)$$

If we set  $\alpha_{js} = p_s$  and  $\beta_{js} = 0$  we get the  $L_{R-I}$  algorithm mentioned earlier. This is the algorithm considered in [14] for the Game problem. The algorithm is easily seen to satisfy conditions (11).

Absolutely expedient algorithms have been shown to be  $\varepsilon$ -optimal with respect to the learning parameter  $b$  [11]. The first conditions for absolute expediency were given in [10]. In these class of algorithms,  $\alpha_{js}(p) = \alpha_s(p)$  and  $\beta_{js}(p) = \beta_s(p)$ . Necessary and sufficient conditions for these class of algorithms to be absolutely expedient are

$$\begin{aligned}\alpha_j/p_j &= \lambda(p) \quad \forall j \\ \beta_j/p_j &= \mu(p) \quad \forall j\end{aligned}\tag{12}$$

A set of conditions which are easily seen to be more general than the (12) conditions but more restrictive than the (11) conditions are

$$\begin{aligned}p_s \alpha_{sj} &= p_j \alpha_{js} \quad \forall j, s \quad s \neq j \\ p_s \beta_{sj} &= p_j \beta_{js} \quad \forall j, s \quad s \neq j\end{aligned}\tag{13}$$

$\alpha$  and  $\beta$  which satisfy (13) but not (12) are

$$\alpha_{js} = \beta_{js} = p_j p_s^2 \quad \forall j, s, \quad s \neq j\tag{14}$$

The following simple lemma will be needed in the the next section.

*Lemma 3.1* If  $p_j = 0$  or  $p_s = 0$ , and the Aso-Kimura conditions (11) are satisfied, then

$$p_s(\alpha_{sj} + \beta_{sj}) = 0\tag{15}$$

*Proof.* Trivially, condition (15) is satisfied if  $p_s = 0$ . Let  $p_s \neq 0$  and  $p_j = 0$ . As  $\alpha_{sj} \leq p_j$ ,  $\alpha_{sj} = 0$ . By the Aso-Kimura conditions (11),

$$\begin{aligned}\sum_{a \neq j} p_a \beta_{aj} &= \sum_{a \neq j} p_j \beta_{ja} \\ &= 0 \text{ as } p_j = 0.\end{aligned}$$

Thus  $\sum_{a \neq j} p_a \beta_{aj} = 0$  and as each term is non-negative,  $p_a \beta_{aj} = 0$  for all  $a$ . In particular,  $p_s \beta_{sj} = 0$ . ■

It is also seen from the above proof that

$$p_j = 0 \Rightarrow (\alpha_{sj} + \beta_{sj}) = 0.\tag{16}$$

An additional condition which will be used in certain proofs in the next section is the following

$$p_j \neq 0 \Rightarrow (\alpha_{sj} + \beta_{sj}) > 0.\tag{17}$$

This is satisfied, for example, by the  $L_{R-I}$  algorithm.

We consider an  $N$ -person game where the  $i$ th player has  $m_i$  pure strategies. We will represent each player by a learning automaton and the actions of the automaton are the pure strategies of the player. Let  $\mathbf{p}_i(k) = [p_{i1}(k) \cdots p_{im_i}(k)]'$  denote the action probability distribution of the  $i$ th player.  $p_{ij}(k)$  denotes the probability with which  $i$ th automaton player chooses the  $j$ th pure strategy at instant  $k$ . Thus  $\mathbf{p}_i(k)$  is the strategy employed by the  $i$ th player at instant  $k$ . Each play of the game consists of each of the automata players choosing an action independently and at random according to their current action probabilities. The payoff to the  $i$ th player will be the reaction to the  $i$ th automaton which will be denoted by  $r_i(k)$ . The learning algorithm used by each of the player is as given below.

1. At each instant  $k$ , player  $i$  selects an action from his action set  $S_i$  according to his current action probability vector  $\mathbf{p}_i(k)$ . Thus, if  $a_i(k)$  is the action at instant  $k$  and  $S_i = \{a_{i1}, \dots, a_{im_i}\}$ , then

$$\text{Prob}(a_i(k) = a_{ij}) = p_{ij}(k).$$

2. Based on the actions taken by all the players, player  $i$  receives a payoff  $r_i(k)$  given by (1).
3. Let the action of the  $i$ th player at instant  $k$  be  $a_{ij}$ . Every player updates his action probabilities according to the following rule.

$$\begin{aligned} p_{is}(k+1) &= p_{is}(k) - br_i(k)\alpha_{ijs}(\mathbf{p}_i(k)) + b(1 - r_i(k))\beta_{ijs}(\mathbf{p}_i(k)) \quad s \neq j \\ p_{ij}(k+1) &= p_{ij}(k) + br_i(k) \sum_{s \neq j} \alpha_{ijs}(\mathbf{p}_i(k)) - b(1 - r_i(k)) \sum_{s \neq j} \beta_{ijs}(\mathbf{p}_i(k)) \end{aligned} \quad (18)$$

where  $0 < b < 1$  is the learning parameter. For simplicity of notation, it is assumed here that all players use the same value of  $b$ . All the results would hold even if different values of the learning parameter were used by the players.  $\alpha_{ijs}$ ,  $\beta_{ijs}$  are the functions used by player  $i$  to update his strategy and are analogous to  $\alpha_{js}$  and  $\beta_{js}$  functions for the single automaton described in the previous section. It may be noted that different players may use different functions to update their strategies provided the functions satisfy the conditions described below.

The  $\alpha$  and  $\beta$  functions satisfy the Aso-Kimura conditions for absolute expediency. They are

$$\begin{aligned} \sum_{s \neq j} p_{is} \alpha_{ijs} &= \sum_{s \neq j} p_{ij} \alpha_{ijs} \\ \sum_{s \neq j} p_{is} \beta_{ijs} &= \sum_{s \neq j} p_{ij} \beta_{ijs} \end{aligned} \quad (19)$$

$\alpha_{ijs}$  and  $\beta_{ijs}$  are non-negative and satisfy conditions (10) to keep  $\mathbf{p}_i(k+1)$  an action probability vector. They can be written as

$$\begin{aligned} \alpha_{ijs}(\mathbf{p}_i) &\leq p_{is} \quad \forall i, j, s, \quad s \neq j \\ \sum_{s \neq j} \beta_{ijs}(\mathbf{p}_i) &\leq p_{ij} \quad \forall i, j \end{aligned} \quad (20)$$

In addition to these conditions an additional condition is imposed to ensure that the updating is not stopped prematurely. This is

$$\alpha_{ijs} + \beta_{ijs} \neq 0 \text{ if } p_{ij} \neq 0 \text{ and } p_{is} \neq 0. \quad (21)$$

#### 4.1 Analysis of the algorithm

The analysis of the algorithm is carried out in two stages. First, weak convergence techniques are used to show that the algorithm can be approximated by an appropriate ODE (Ordinary Differential Equation) as  $b \rightarrow 0$ . Then, solutions of the ODE are analysed to obtain information about the behaviour of the algorithm.

The learning algorithm given by (18) can be represented as

$$P(k+1) = P(k) + bG(P(k), a(k), r(k)) \quad (22)$$

where  $a(k) = (a_1(k) \cdots a_N(k))$  denotes the actions selected by the automata team at  $k$  and  $r(k) = (r_1(k) \cdots r_N(k))$  are the resulting payoffs.

Let  $P(k) = (\mathbf{p}_1(k), \dots, \mathbf{p}_N(k))$  denote the vector current mixed strategies of all the players which is also the state of the learning algorithm at instant  $k$ . Our interest is in the asymptotic behaviour of  $P(k)$ . Since each  $\mathbf{p}_i$  is a probability vector, we have  $P(k) \in \mathbf{K}$  where  $\mathbf{K}$  is as defined by (5). The following piecewise-constant interpolation of  $P(k)$  is required to use the weak convergence techniques

$$P^b(t) = P(k), \quad t \in [kb, (k+1)b) \quad (23)$$

$P^b(\cdot) \in D^{m_1 + \cdots + m_N}$ , the space of all functions from  $\mathbb{R}$  into  $[0, 1]^{m_1 + \cdots + m_N}$ , which are continuous on the right and have limits on the left. (It may be noted that  $P^b(t) \in \mathbf{K}, \forall t$ ). Now consider the sequence  $\{P^b(\cdot), b > 0\}$ . We are interested in the limit of this sequence as  $b \rightarrow 0$ .

Define a function  $\xi: \mathbf{K} \rightarrow [0, 1]^{m_1 + \cdots + m_N}$  by

$$\xi(P) = E[G(P(k), a(k), r(k)) | P(k) = P] \quad (24)$$

**Theorem 4.1.** *The sequence of interpolated processes  $\{P^b(\cdot)\}$  converges weakly, as  $b \rightarrow 0$ , to  $X(\cdot)$  which is the solution of the ODE,*

$$\frac{dX}{dt} = \xi(X), \quad X(0) = P_0 \quad (25)$$

*Proof.* The following conditions are satisfied by the learning algorithm given by (22)

1.  $\{P(k), (a(k-1), r(k-1)), k \geq 0\}$  is a Markov process.  $(a(k), r(k))$  take values in a compact metric space.
2. The function  $G(\cdot, \cdot, \cdot)$  is bounded and continuous and independent of  $b$ .
3. If  $P(k) \equiv P$ , then  $\{(a(k), r(k)), k \geq 0\}$  is an i.i.d. sequence. Let  $M^P$  denote the (invariant) distribution of this process.
4. The ODE (25) has a unique solution for each initial condition.

Hence by [8, Thm. 3.2], the sequence  $\{P^b(\cdot)\}$  converges weakly as  $b \rightarrow 0$  to the solution of the ODE,

$$\frac{dX}{dt} = G(X), X(0) = P_0$$

where  $\bar{G}(P) = E^P G(P(k), a(k), r(k))$  and  $E^P$  denotes the expectation with respect to the invariant measure  $M^P$ .

Since for  $P(k) \equiv P$ ,  $(a(k), r(k))$  is i.i.d. whose distribution depends only on  $P$  and the payoff matrices,

$$\bar{G}(P) = E[G(P(k), a(k), r(k)) | P(k) = P] = \xi(P), \quad \text{by (24)}$$

Hence the theorem. ■

**Remark 4.1.** The convergence of functionals implied by weak convergence, along with the knowledge of the nature of the solutions of the ODE (25), enables one to understand the long-term behaviour of  $P(k)$ . It can be shown that  $P(k)$  follows the trajectory  $X(t)$  of the ODE within a prespecified error and for as long as prespecified with probability increasingly close to 1 as  $b$  decreases. (See [3, Chapter 2, Theorem 1] and the discussion therein).

The following lemma gives an explicit characterisation of  $\xi$ .

**Lemma 4.1**

$$\xi_{ij}(P) = \sum_{s \neq j} p_{is} (\alpha_{isj} + \beta_{isj}) (h_{ij} - h_{is}) \quad (26)$$

where  $h_{is}$  are as defined by (7).

**Proof.** Let  $G_{ij}$  denote the  $(i, j)$ th component of  $G$ .

$$\begin{aligned} \xi_{ij}(P) &= E[G_{ij}(P(k), a(k), r(k)) | P(k) = P] \\ &= \sum_s E[G_{ij}(P(k), a(k), r(k)) | P(k) = P, a_i(k) = a_{is}] p_{is} \\ &= \sum_{s \neq j} E[r_i(k) \alpha_{ijs}(\mathbf{p}_i(k)) - (1 - r_i(k)) \beta_{ijs}(\mathbf{p}_i(k)) | P(k) = P, a_i(k) = a_{ij}] p_{ij} \\ &\quad + \sum_{s \neq j} E[-r_i(k) \alpha_{isj}(\mathbf{p}_i(k)) + (1 - r_i(k)) \beta_{isj}(\mathbf{p}_i(k)) | P(k) \\ &\quad = P, a_i(k) = a_{is}] p_{is} \\ &= \sum_{s \neq j} \{ p_{ij} \alpha_{ijs}(\mathbf{p}_i) E[r_i | P, a_{ij}] - p_{ij} \beta_{ijs}(\mathbf{p}_i) E[1 - r_i | P, a_{ij}] p_{ij} \} \\ &\quad + \sum_{s \neq j} \{ -p_{is} \alpha_{isj}(\mathbf{p}_i) E[r_i | P, a_{is}] - p_{is} \beta_{isj}(\mathbf{p}_i) E[1 - r_i | P, a_{is}] p_{is} \} \\ &= \sum_{s \neq j} p_{ij} (\alpha_{ijs} h_{ij} - \beta_{ijs} (1 - h_{ij})) + \sum_{s \neq j} -p_{is} (\alpha_{isj} h_{is} - \beta_{isj} (1 - h_{is})) \\ &= \sum_{s \neq j} p_{is} (\alpha_{isj} + \beta_{isj}) (h_{ij} - h_{is}) \text{ by the Aso-Kimura conditions (19)} \end{aligned}$$

completing the proof. ■

Using (26), the ODE (25) can be written as

$$\frac{dp_{ij}}{dt} = \sum_{s \neq j} p_{is} (\alpha_{isj}(\mathbf{p}_i) + \beta_{isj}(\mathbf{p}_i)) [h_{ij}(P) - h_{is}(P)] \quad 1 \leq j \leq m_i, \quad 1 \leq i \leq N. \quad (27)$$

The following theorem characterises the solutions of the ODE and hence the long-term behaviour of the algorithm.

**Theorem 4.2** *The following are true of the ODE (and hence of the learning algorithm if the parameter  $b$  in (18) is sufficiently small).*

1. All corners of  $\mathbf{K}$  (i.e. points in  $\mathbf{K}^*$ ) are stationary points.
2. All Nash equilibria are stationary points.
3. If conditions (13) and (17) are satisfied, all stationary points that are not Nash equilibria are unstable.
4. All corners of  $\mathbf{K}$  that are strict Nash equilibria are locally asymptotically stable.

*Proof.* 1. Let  $P^0 \in \mathbf{K}^*$ . Thus  $p_{ij}^0 = 0$  or  $p_{is}^0 = 0$  for all  $j, s$  such that  $j \neq s$  and  $1 \leq j, s \leq m_i$ .

By Lemma 3.1,  $p_{is}^0(\alpha_{isj} + \beta_{isj}) = 0$  for all  $s$ . Thus,

$$\frac{dp_{ij}}{dt} = \sum_{s \neq j} p_{is} (\alpha_{isj}(\mathbf{p}_i^0) + \beta_{isj}(\mathbf{p}_i^0)) [h_{is}(P^0) - h_{ij}(P^0)] = 0$$

2. Let  $P^0$  be a Nash equilibrium. Define

$$A_i = \{s: p_{is}^0 > 0\}$$

For  $j \notin A_i$ ,  $p_{ij}^0 = 0$ . Thus,  $p_{is}^0(\alpha_{isj} + \beta_{isj}) = 0$  for all  $s \neq j$ .

Let  $j \in A_i$ , implying  $p_{ij}^0 > 0$ . It is trivially seen that  $p_{is}^0(\alpha_{isj} + \beta_{isj})(h_{ij} - h_{is}) = 0$  if  $p_{is}^0 = 0$ . Therefore let  $p_{is}^0 > 0$ . But then as  $P^0$  is a Nash equilibrium,  $h_{ij}(P^0) = h_{is}(P^0)$  by Lemma 2.1. Thus

$$p_{is}^0(\alpha_{isj} + \beta_{isj})(h_{ij} - h_{is}) = 0 \quad \forall i, j, s \quad (s \neq j)$$

Thus  $P^0$  is a stationary point.

3. Let  $P^0$  be a zero of  $\xi(\cdot)$  which is not a Nash equilibrium. It is assumed that conditions (13) and (17) are satisfied by the algorithm. By Lemma 2.1, there is an  $i$  and an  $s$  such that

$$h_{is}(P^0) > g^i(P^0) \quad (28)$$

In general, there will be more than one  $s$  such that  $h_{is}(P^0) > g^i(P^0)$ . Without loss of generality let

$$h_{i1}(P^0) = h_{i2}(P^0) = \dots = h_{iL}(P^0) > h_{iL+1}(P^0) \geq h_{iL+2}(P^0) \geq \dots$$

where  $h_{is}(P^0) > g^i(P^0)$ ,  $1 \leq s \leq L$ . Then, for all  $\delta, \varepsilon$  sufficiently small, there exists a neighbourhood  $\mathcal{U}_{\varepsilon\delta}$  around  $P^0$  such that for all  $P \in \mathcal{U}_{\varepsilon\delta}$ ,  $h_{ij}(P) - h_{is}(P) > \varepsilon$  for all  $j \leq L$ ,  $s > L$  and  $h_{ij}(P) - h_{is}(P) > -\delta\varepsilon$  for all  $s, j \leq L$ . Then



$$\begin{aligned}
\frac{dp_{i1}}{dt} &= \sum_{s \neq 1} p_{is} (\alpha_{is1} + \beta_{is1}) (h_{i1} - h_{is}) \\
&= p_{i1} \sum_{s \neq 1} (\alpha_{i1s} + \beta_{i1s}) (h_{i1} - h_{is}) \text{ as (13) is satisfied} \\
&= p_{i1} \sum_{s > L} (\alpha_{i1s} + \beta_{i1s}) (h_{i1} - h_{is}) + p_{i1} \sum_{2 \leq s \leq L} (\alpha_{i1s} + \beta_{i1s}) (h_{i1} - h_{is}) \\
&> \varepsilon p_{i1} \left( \sum_{s > L} (\alpha_{i1s} + \beta_{i1s}) - \delta \sum_{2 \leq s \leq L} (\alpha_{i1s} + \beta_{i1s}) \right) \text{ if } P \in \mathcal{U}_{\varepsilon\delta}
\end{aligned}$$

There is at least one  $s > L$  such that  $p_{is}^0 > 0$ , as  $P^0$  is not a Nash equilibrium. Thus,  $\alpha_{i1s}(P^0) + \beta_{i1s}(P^0) > 0$ , by (17). Thus the linearised version of the last line in the above equation is strictly positive and thus  $P^0$  is not stable.

It should be possible to relax condition (17), but the analysis would then involve higher order terms instead of just linear terms. But the same result should go through.

4. Let  $P^0 = (e_{i_1}, \dots, e_{i_N})$  be a corner of  $\mathbf{K}$  that is a Nash equilibrium. Without loss of generality, let  $i_1 = i_2 = \dots = i_N = 1$ . Use the transformation  $P \rightarrow \varepsilon$ , defined by

$$\begin{aligned}
\varepsilon_{iq} &= p_{iq} \text{ if } q \neq 1 \\
\varepsilon_{i1} &= 1 - p_{i1}
\end{aligned}$$

Define a Lyapunov function  $V(\cdot)$  as

$$V(\varepsilon) = \sum_{i=1}^N \varepsilon_{i1}$$

It is easy to see that  $V \geq 0$  and that  $V=0$  iff  $\varepsilon=0$ . Also, as  $P^0$  is a strict Nash equilibrium,  $h_{i1}(P^0) > h_{is}(P^0)$  for all  $s$ . Thus,

$$\begin{aligned}
\frac{dV}{dt} &= \sum_{i=1}^N \frac{d\varepsilon_{i1}}{dt} \\
&= - \sum_{i=1}^N \frac{dp_{i1}}{dt} \\
&= - \sum_{i=1}^N \sum_{s \neq 1} p_{is} (\alpha_{is1} + \beta_{is1}) (h_{i1} - h_{is}) \\
&= - \sum_{i=1}^N \sum_{s \neq 1} \varepsilon_{is} (\alpha_{is1}(\mathbf{p}_i^0) + \beta_{is1}(\mathbf{p}_i^0)) (h_{i1}(P^0) - h_{is}(P^0)) \\
&\quad + \text{higher order terms in } \varepsilon \\
&< 0 \text{ as } h_{i1}(P^0) > h_{is}(P^0) \quad \forall s \neq 1
\end{aligned}$$

This proves  $P^0$  is asymptotically stable. ■

*Remark 4.2.* Because of the above theorem, we can conclude that our learning algorithm will never converge to a point in  $\mathbf{K}$  which is not a Nash equilibrium and

strict Nash equilibria in pure strategies are locally asymptotically stable. This still leaves two questions unanswered. (i) Do Nash equilibria in mixed strategies form stable attractors for the algorithm, and (ii) is it possible that  $P(k)$  does not converge to a point in  $\mathbf{K}$  which would be the case, for example, if the algorithm exhibits limit cycle behaviour. At present we have no results concerning the first question. Regarding the second question we provide a sufficient condition for  $P(k)$  to converge to some point in  $\mathbf{K}$ . This is proved in Theorem 4.3 below.

**Theorem 4.3.** *Suppose there is a bounded differentiable function*

$$F: \mathbb{R}^{m_1 + \dots + m_N} \rightarrow \mathbb{R}$$

*such that for some constants  $c_i > 0$ ,*

$$\frac{\partial F}{\partial p_{iq}}(P) = c_i h_{iq}(P), \quad \forall i, q \text{ and all } P \in \mathbf{K}. \quad (29)$$

*Further, the  $\alpha$  and  $\beta$  functions satisfy condition (13) and (17). Then the learning algorithm, for any initial condition in  $\mathbf{K} - \mathbf{K}^*$ , always converges to a Nash equilibrium.*

*Proof.* Consider the variation of  $F$  along the trajectories of the ODE. We have

$$\begin{aligned} \frac{dF}{dt} &= \sum_{i,q} \frac{\partial F}{\partial p_{iq}} \frac{dp_{iq}}{dt} \\ &= \sum_{i,q} \frac{\partial F}{\partial p_{iq}}(P) \sum_s p_{is} (\alpha_{isq} + \beta_{isq}) [h_{iq}(P) - h_{is}(P)], \quad \text{by (26)} \\ &= \sum_i c_i \sum_q \sum_s p_{is} (\alpha_{isq} + \beta_{isq}) [(h_{iq}(P))^2 - h_{iq}(P)h_{is}(P)], \quad \text{by (29)} \\ &= \sum_i c_i \sum_q \sum_{s>q} p_{is} (\alpha_{isq} + \beta_{isq}) [h_{iq}(P) - h_{is}(P)]^2 \\ &\geq 0 \end{aligned} \quad (30)$$

Thus  $F$  is nondecreasing along the trajectories of the ODE. Also, due to the nature of the learning algorithm given by (18), all solutions of the ODE (25), for initial conditions in  $\mathbf{K}$ , will be confined to  $\mathbf{K}$  which is a compact subset of  $\mathbb{R}^{m_1 + \dots + m_N}$ . Hence by [12, Theorem 2.7], asymptotically all the trajectories will be in the set  $\mathbf{K}_1 = \{P \in [0, 1]^{m_1 + \dots + m_N} : dF/dt(P) = 0\}$ .

From (30), (25) and (26) it is easy to see that

$$\frac{dF}{dt}(P) = 0 \Rightarrow p_{is} (\alpha_{isq} + \beta_{isq}) [h_{iq}(P) - h_{is}(P)] = 0 \quad \forall q, s, i$$

$$\Rightarrow \xi_{iq}(P) = 0 \quad \forall i, q$$

$$\Rightarrow P \text{ is a stationary point of the ODE}$$

Theorem 4.2, and Theorem 4.3 together characterise the long-term behaviour of the learning algorithm. For any general  $N$ -person game, all strict Nash equilibria in pure strategies are asymptotically stable in the small. Further the algorithm cannot converge to any point in  $\mathbf{K}$  which is not a Nash equilibrium. If the game satisfies the sufficiency condition needed for Theorem 4.3 then the algorithm will converge to a Nash equilibrium. (If the game does not satisfy this condition we cannot be sure that the algorithm will converge rather than, e.g., exhibit a limit cycle behaviour). We have not been able to establish that, in a general game, all mixed strategy equilibria are stable attractors.

## 5. Discussion

In this paper we have considered an  $N$ -person stochastic game with incomplete information. We presented a method based on Learning Automata for the players to learn equilibrium strategies in a decentralised fashion. In our framework, each player can choose his own learning algorithm. However, the algorithm of each player should satisfy the so called absolute expediency property. In the context of a learning automation, an algorithm is absolutely expedient if the expected reinforcement will increase monotonically with time in all stationary random environments. In the context of the Game, since all the players are updating their strategies, the effective environment as seen by a player will be nonstationary. Here, the restriction of absolute expediency will mean that the algorithm used by a player should ensure the expected payoff to the player increases monotonically when all other players are using fixed strategies. Thus, this is a mild requirement because rational behaviour on the part of the player demands that, at the minimum, he should strive to improve his payoff when everyone else is playing to a fixed strategy. The analysis presented in §4 tells us that if all players are using absolutely expedient algorithms, they can converge to Nash equilibria without needing any information exchange.

In a truly competitive game situation, we may require that the learning algorithm employed by a player should also help to confuse the opponent. The framework presented in this paper does not address this aspect. The analysis presented here also does not address the question of how a player, using an absolutely expedient algorithm, will perform if other players are using arbitrary learning algorithms that are not absolutely expedient. However, in many cases, game models are employed as techniques for solving certain type of optimisation problems. Examples include learning optimal discriminant functions in Pattern Recognition and the Consistent Labelling problem in Computer Vision [17, 18]. (See [14] for a discussion on the application of the Game model considered here in such problems. In such applications the sufficiency condition of Theorem 4.3 also holds). In all such situations, the requirement that the players use absolutely expedient learning algorithms is not restrictive.

Our algorithm can be used for learning the Nash equilibrium even if the game is deterministic and the game matrix is known. We then simply make  $r_i$ , payoff to the  $i$ th player, equal to (suitably normalised)  $d^i(a_1, \dots, a_N)$  which is the game matrix entry corresponding to the actions played. Whether this is an efficient algorithm for obtaining Nash equilibria for general deterministic games with known payoff functions

Our analysis does not establish the stability or otherwise of Nash equilibria in mixed strategies for the general  $N$ -person game. In the context of a single Learning Automaton, it is known that in any stationary random environment absolutely expedient algorithms always converge to a unit vector [13] and hence it might be the case that even the decentralised team cannot converge to an interior point. This aspect needs further investigation.

## Acknowledgement

This work was supported by the Office of Naval Research Grant No. 00014-92-J-1324 under an Indo-US project.

## References

- [1] Aso H and Kimura M, Absolute expediency of learning automata. *Inf. Sci.* **17** (1976) 91–122
- [2] Basar T and Olsder G J, *Dynamic noncooperative game theory*, (New York: Academic Press) (1982)
- [3] Albert Beneveniste, Michel Metivier and Pierre Priouret, *Adaptive algorithms and stochastic approximations*. (New York: Springer Verlag) (1987)
- [4] Binmore, *Essays on foundations of game theory*. (Basil Blackwell) (1990)
- [5] Friedman J W, *Oligopoly and the theory of games*. (New York: North Holland) (1977)
- [6] Harsanyi J C, Games with incomplete information played by bayesian player – I, *Manage. Sci.* **14** (1967) 159–182
- [7] Wang Jianhua, *The theory of games*. (Oxford: Clarendon Press) (1988)
- [8] Kushner H J, *Approximation and weak convergence methods for random processes* (Cambridge MA: MIT Press) (1984)
- [9] Lakshmivarahan S and Narendra K S, Learning algorithms for two-person zero-sum stochastic games with incomplete information: a unified approach, *SIAM J. Control Optim.* **20** (1982) 541–552
- [10] Lakshmivarahan S and Thathachar M A L, Absolutely expedient learning algorithms for stochastic automata. *IEEE Trans. Syst., Man Cybern.* **3** (1973) 281–286
- [11] Meybodi M R and Lakshmivarahan S,  $\epsilon$ -optimality of a general class of absorbing barrier learning algorithms, *Inf. Sci.* **28** (1982) 1–20
- [12] Narendra K S and Annaswamy A, *Stable adaptive systems*, (Englewood Cliffs: Prentice Hall) (1989)
- [13] Narendra K S and Thathachar M A L, *Learning automata: An introduction*. (Englewood Cliffs: Prentice Hall) (1989)
- [14] Sastry P S, Phansalkar V V and Thathachar M A L, Decentralized learning of Nash equilibria in multi-person stochastic games with incomplete information. To appear in *IEEE Tran. Syst., Man Cybern.*
- [15] Srikanta Kumar P R and Narendra K S, A learning model for routing in telephone networks, *SIAM J. Control Optim.* **20** (1982) 34–57
- [16] Thathachar M A L and Ramakrishnan K R, A cooperative game of a pair of automata. *Automatica* **20** (1984) 797–801
- [17] Thathachar M A L and Sastry P S, Learning optimal discriminant functions through a cooperative game of automata, *IEEE Trans. Syst., Man Cybern.* **17**(1) (1957) 73–85
- [18] Thathachar M A L and Sastry P S, Relaxation labelling with learning automata, *IEEE Trans. Pattern Anal. Mach. Intelligence*, **8** (1986) 256–268
- [19] Vishwanatha Rao T, *Learning solutions to stochastic non-cooperative games*, ME Thesis, (Dept. of Electrical Engineering, Indian Institute of Science), Bangalore, India (1984)
- [20] Wheeler Jr R M, and Narendra K S, Decentralized learning in finite markov chains, *IEEE Trans. Autom. Control*. **31**(6) (1986) 519–526

## Hierarchic control

J L LIONS

Collège de France, 3 Rue D'ulm; 75231 Paris Cedex 05, France

Dedicated to the memory of Professor K G Ramanathan

**Abstract.** Distributed control is applied to a system modelled by a parabolic evolution equation. One considers situations where there are two cost (objective) functions. One possible way is to cut the control into 2 parts, one being thought of as “the leader” and the other one as “the follower”. This situation is studied in the paper, with one of the cost functions being of the controllability type. Existence and uniqueness is proven. The optimality system is given in the paper.

**Keywords.** Hierarchic control; stackleberg terminology; optimality system.

### 1. Introduction

Let us firstly recall briefly what is meant by *controllability* for distributed systems, in the case of the *wave equation*.

We shall then introduce what we mean by “*hierarchic control*” in the framework of controllability. □

Let  $\Omega$  be a bounded open set of  $\mathbb{R}^n$ , with smooth boundary  $\Gamma$ .

Let  $\Gamma_0$  be a subset of  $\Gamma$ . *The control will be applied on  $\Gamma_0$ , and will depend on  $t$ .*

The state  $y$  of the system is given by

$$\frac{\partial^2 y}{\partial t^2} - \Delta y = 0 \text{ in } \Omega \times (0, T) \quad (1.1)$$

subject to

$$y = \begin{cases} v & \text{on } \Sigma_0 = \Gamma_0 \times (0, T) \\ 0 & \text{on } \Sigma \setminus \Sigma_0, \Sigma = \Gamma \times (0, T) \end{cases} \quad (1.2)$$

and with the initial conditions

$$y|_{t=0} = \frac{\partial y}{\partial t} \Big|_{t=0} = 0 \text{ in } \Omega. \quad (1.3)$$

In order to make things precise, let us assume that

$$v \in L^2(\Sigma_0). \quad (1.4)$$

Then, (1.1), (1.2), (1.3) admit a unique solution, denoted by

$$\begin{cases} y(x, t; v) = y(v) \\ y(v) \in L^2(\Omega \times (0, T)). \end{cases} \quad (1.5)$$

For the proof of (1.5), one can consult J L Lions [4], Chapter 2, Sections 4.2, 4.4 and [5].

One can prove that – may be after modifying the solution on a set of measure 0 on  $(0, T)$ –

$$\left| \begin{array}{l} t \rightarrow \left\{ y(t; v), \frac{\partial y}{\partial t}(t; v) \right\} \text{ is continuous} \\ \text{from } [0, T] \rightarrow L^2(\Omega) \times H^{-1}(\Omega) \end{array} \right. \quad (1.6)$$

where

$$H^{-1}(\Omega) = \text{dual of } H_0^1(\Omega),$$

$$H_0^1(\Omega) = \left\{ \varphi \mid \varphi, \frac{\partial \varphi}{\partial x_1}, \dots, \frac{\partial \varphi}{\partial x_n} \in L^2(\Omega), \varphi = 0 \text{ on } \Gamma \right\}$$

and where  $y(t; v)$  denotes the function  $x \rightarrow y(x, t; v)$ .

The problem of *exact controllability* is the following: let  $y^0, y^1$  be given arbitrarily in  $L^2(\Omega) \times H^{-1}(\Omega)$ , and let  $T > 0$  be given.

One wants to find  $v \in L^2(\Sigma_0)$  such that

$$y(T; v) = y^0, \quad \frac{\partial y}{\partial t}(T; v) = y^1. \quad (1.7)$$

One says that one has *exact controllability* if such a control  $v$  exists for any couple  $\{y^0, y^1\}$ . In this case, there are infinitely many  $v$ 's giving (1.7). One reasonable choice is then to consider

$$\inf \frac{1}{2} \int_{\Sigma_0} v^2 d\Sigma, \quad v \text{ subject to (1.7)}. \quad (1.8)$$

For a study of this problem we refer to J L Lions [5], Vol. 1. □

*Approximate controllability* comes next. Let  $B_0$  (resp.  $B_{-1}$ ) be the unit ball of  $L^2(\Omega)$  (resp.  $H^{-1}(\Omega)$ ) and let  $\alpha_0, \alpha_1$  be arbitrarily small positive numbers. We “relax” (1.7) by

$$y(T; v) \in y^0 + \alpha_0 B_0, \quad \frac{\partial y}{\partial t}(T; v) \in y^1 + \alpha_1 B_{-1}, \quad (1.9)$$

If there is a control  $v$  satisfying (1.8) for every couple  $\{y^0, y^1\}$  and for every  $\alpha_0, \alpha_1 > 0$ , one says that one has *approximate controllability*.

Of course approximate controllability is equivalent to

$$\left\{ y(T; v), \frac{\partial y}{\partial t}(T; v) \right\} \text{ spans a dense subset of } L^2(\Omega) \times H^{-1}(\Omega). \quad (1.9 \text{ bis})$$

One can prove (cf. J L Lions [5], Vol. 1), in particular using Holmgren's uniqueness theorem (cf. loc. cit., Th. 8.2, Chapter 1), that

$$\left| \begin{array}{l} \text{approximate controllability holds true if (and only if)} \\ T > 2d(\Omega, \Gamma_0), \end{array} \right. \quad (1.10)$$

where

$$\begin{cases} d(\Omega, \Gamma_0) = \sup_{x \in \Omega} d(x, \Gamma_0), \\ d(x, \Gamma_0) = \text{distance from } x \text{ to } \Gamma_0 \text{ taken in the geodesic sense in } \Omega. \end{cases} \quad (1.11)$$

In this case, we are looking for

$$\inf \frac{1}{2} \int_{\Sigma_0} v^2 d\Sigma, \quad v \text{ subject to (1.9).} \quad (1.12)$$

We refer to J L Lions [5] for this question, where, in particular, we derive the *optimality system*, i.e. the set of P.D.E's (Partial Differential Equation) and of Variational Inequalities which characterize the unique solution of (1.12).  $\square$

But there are many situations—it is almost always the case!—where achieving (1.8) is *not the only criteria*.

For instance—and this will be the question we shall address here—one may want to achieve (1.8) but we also want that, during the whole interval  $t \in (0, T)$ ,  $y(t; v)$  “does not go too far” from a given state  $y_2(x, t)$ . In a quantitative manner, we introduce

$$\mathcal{J}(v) = \frac{1}{2} \int \int_{\Omega \times (0, T)} (y(x, t; v) - y_2(x, t))^2 dx dt + \frac{\beta}{2} \int_{\Sigma_0} v^2 d\Sigma \quad (1.13)$$

and we would like to “minimize” (1.13).

But of course minimizing at the same time (1.12) and (1.13) does not make sense.  $\square$

This is a question of *multicriteria optimization*—a situation which is classical in economy. Precisely possible notions that one can use when dealing with these questions are coming from Economy with Pareto optimal control and with Stackleberg [10] Optimization.

We divide  $\Gamma_0$  in two parts  $\Gamma_1, \Gamma_2 \subset \Gamma$

$$\Gamma_0 = \Gamma_1 \cup \Gamma_2 \quad (1.14)$$

and we consider

$$v = \{v_1, v_2\}, \quad v_i = \text{control function in } L^2(\Sigma_i), \quad i = 1, 2. \quad (1.15)$$

We can also write

$$v = v_1 + v_2 \quad (1.16)$$

with

$$\Gamma_1 = \Gamma_2 = \Gamma_0.$$

We assume that there is a *hierarchy* in our wishes. The main objective is to have (1.8) at “minimum cost”. The second priority is to maintain (1.13) as small as possible. In the decomposition (1.14), (1.15) we also establish a hierarchy. We think of  $v_1$  as being the “main” control, *the leader* (in Stackleberg terminology), and we think of  $v_2$  as *the follower*, always in Stackleberg terminology.

Let us set

$$y(v) = y(v_1, v_2). \quad (1.16)$$

We write (1.13) as

$$J_2(v_1, v_2) = \frac{1}{2} \int \int_{\Omega \times (0, T)} (y(v_1, v_2) - y_2)^2 dx dt + \frac{\beta}{2} \int_{\Sigma_2} v_2^2 d\Sigma \quad (1.17)$$

(we drop in the last term  $\frac{\beta}{2} \int_{\Sigma_1} v_1^2 d\Sigma$ , without changing anything as it will be clear below).

Then given  $v_1$  we consider

$$\inf_{v_2 \in L^2(\Sigma_2)} J_2(v_1, v_2). \quad (1.18)$$

This is a classical type problem in the control of distributed systems (cf. J L Lions [8]). It admits a unique solution

$$v_2 = \mathcal{F}(v_1) \quad (1.19)$$

given by an optimality system that we present in § 2 below.

We then consider the state

$$y(v_1, \mathcal{F}(v_1)) \quad (1.20)$$

given by the solution of (compare to (1.1), (1.2), (1.3)):

$$\begin{aligned} \frac{\partial^2 y}{\partial t^2} - \Delta y &= 0 \text{ in } \Omega \times (0, T), \\ y &= \begin{cases} v_1 \text{ on } \Sigma_1 = \Gamma_1 \times (0, T) \\ \mathcal{F}(v_1) \text{ on } \Sigma_2 = \Gamma_2 \times (0, T) \\ 0 \text{ on } \Sigma \setminus \Sigma_1 \cup \Sigma_2 = \Sigma \setminus \Sigma_0, \end{cases} \\ y|_{t=0} = \frac{\partial y}{\partial t} \Big|_{t=0} &= 0 \text{ in } \Omega. \end{aligned} \quad (1.21)$$

Then we address the main criteria, i.e. we want to find

$$\inf \frac{1}{2} \int_{\Sigma_1} v_1^2 d\Sigma, \quad (1.22)$$

$v_1$  subject to

$$\begin{aligned} y(T; v_1, \mathcal{F}(v_1)) &\in y^0 + \alpha_0 B_0, \\ \frac{\partial y}{\partial t}(T; v_1, \mathcal{F}(v_1)) &\in y^1 + \alpha_1 B_{-1} \end{aligned} \quad (1.23)$$

provided this is possible  $\forall y^0, y^1, \alpha_0, \alpha_1$ .

Actually we shall prove (§ 3 below) that in case (1.16)

$$\left| \left\{ y(T; v_1, \mathcal{F}(v_1)), \frac{\partial y}{\partial t}(T; v_1, \mathcal{F}(v_1)) \right\} \right| \text{ spans a dense subspace of } L^2(\Omega) \times H^{-1}(\Omega) \quad (1.24)$$



if and only if

$$T > 2d(\Omega, \Gamma_1). \quad (1.25)$$

□

*Remark 1.1.* Since we have

$$d(\Omega, \Gamma_1) \geq d(\Omega, \Gamma_0) \quad (1.26)$$

condition (1.25) is *stronger* than (1.10), as it is natural (since we want to achieve *more*). □

*Remark 1.2.* The idea of dividing  $v$  in two parts  $v_1, v_2$  in order to apply Stackleberg Optimization has been introduced in [7], for systems governed by *parabolic equations*. □

*Remark 1.3.* The fact that  $T$  should be large enough (cf. (1.10), (1.25)) is due to the finite speed of wave propagation. Conditions of this type do not appear for diffusion type systems. □

*Remark 1.4.* The problem addressed in this paper arises from applications. For instance, if we consider the controllability of a *flexible structure* (large space structure, sea-platform, flexible robot, ...), we want to reach a given state at a given time  $T$ , with some constraints in the course of the operations, which *can* be expressed by  $J_2$ . □

*Remark 1.5.* The division of  $v$  in 2 parts  $v_1, v_2$ , can be made in infinitely many ways. It is quite natural to try to maintain  $d(\Omega, \Gamma_1)$  as close as possible from  $d(\Omega, \Gamma_0)$  (cf. (1.25), (1.26)). There are no other *mathematical* considerations. □

*Remark 1.6.* We present the solution in the specific case of the wave equation with Dirichlet's type control. But the method is completely general. Similar considerations could be made for all the situations considered in [5], Vols 1 and 2. □

*Remark 1.7.* One can use similar notions for *non-linear* controllability problems, but with most of the theoretical questions being then *open*. □

*Remark 1.8.* Numerical aspects are not considered here. A survey of numerical methods for controllability problems is given in R Glowinski and the author in [2]. □

## 2. Optimality system for the follower

Let us denote by  $v_2$  the solution of (1.18). We set for a moment

$$y(v_1, v_2) = y.$$

The Euler equation for (1.18) is given by

$$\iint_{\Omega \times (0, T)} (y - y_2) \hat{y} dx dt + \rho_0 \int_{\Sigma_2} v_2 \hat{v}_2 d\Sigma = 0, \quad (2.1)$$

$$\begin{aligned} \hat{y} &= \begin{cases} 0 & \text{on } \Sigma_1, \\ \hat{v}_2 & \text{on } \Sigma_2, \\ 0 & \text{on } \Sigma \setminus \Sigma_1 \cup \Sigma_2 \end{cases} \\ \hat{y}|_{t=0} = \frac{\partial \hat{y}}{\partial t} \Big|_{t=0} &= 0. \end{aligned} \quad (2.2)$$

We introduce  $p$  by

$$\begin{cases} p'' - \Delta p = y - y_2 & \text{in } \Omega \times (0, T), \\ p(T) = p'(T) = 0, p = 0 & \text{on } \Sigma \end{cases} \quad (2.3)$$

where we write from now on,  $\varphi' = \frac{\partial \varphi}{\partial t}$ ,  $\varphi'' = \frac{\partial^2 \varphi}{\partial t^2}$ .

If we multiply (2.3) by  $\hat{y}$  and if we integrate by parts (the integrations by parts being valid) we obtain

$$\int \int_{\Omega \times (0, T)} (y - y_2) \hat{y} dx dt = - \int_{\Sigma_2} \frac{\partial p}{\partial \nu} \hat{v}_2 d\Sigma, \quad (2.4)$$

so that (2.1) becomes

$$\frac{\partial p}{\partial \nu} = \beta v_2 \text{ on } \Sigma_2. \quad (2.5)$$

We can summarize as follows. Given  $v_1$  in  $L^2(\Sigma_1)$ , the follower  $v_2$  is given by

$$v_2 = \mathcal{F}(v_1) = \frac{1}{\beta} \frac{\partial p}{\partial \nu} \text{ on } \Sigma_2 \quad (2.6)$$

where  $\{y, p\}$  is the unique solution of (the optimality system)

$$\begin{aligned} y'' - \Delta y &= 0, \\ p'' - \Delta p &= y - y_2, \\ y(0) = y'(0) &= 0, \quad p(T) = p'(T) = 0, \\ y &= \begin{cases} v_1 & \text{on } \Sigma_1 \\ \frac{1}{\beta} \frac{\partial p}{\partial \nu} & \text{on } \Sigma_2, p = 0 \text{ on } \Sigma, \\ 0 & \text{on } \Sigma \setminus \Sigma_0. \end{cases} \end{aligned} \quad (2.7)$$

Of course,  $\{y, p\}$  depends on  $v_1$ :

$$\{y, p\} = \{y(v_1), p(v_1)\}. \quad (2.8)$$

We have now to find the optimal leader.

We can now rewrite (1.21), (1.22), (1.23) as follows, using (2.7), (2.8). We are looking for

$$\inf \frac{1}{2} \int_{\Sigma_1} v_1^2 d\Sigma \quad (3.1)$$

where  $v_1$  is subject to

$$y(T; v_1) \in y^0 + \alpha_0 B_0, \quad y'(T; v_1) \in y^1 + \alpha_1 B_{-1}, \quad (3.2)$$

assuming that such  $v_1$ 's do exist.  $\square$

We are now going to show that in case (1.16)

$$\left\{ y(T; v_1), y'(T; v_1) \right\} \text{ spans a dense (affine) subspace of } L^2(\Omega) \times H^{-1}(\Omega) \text{ when } v_1 \text{ spans } L^2(\Sigma_1), \quad (3.3)$$

provided that (1.25) holds true.  $\square$

Let us set

$$y = y_0 + z, \quad p = p_0 + q, \quad (3.4)$$

where  $\{y_0, p_0\}$  is given by

$$\begin{aligned} y_0'' - \Delta y_0 &= 0, \\ p_0'' - \Delta p_0 &= y_0 - y_2, \\ y_0(0) = y_0'(0) &= 0, \quad p_0(T) = p_0'(T) = 0 \text{ in } \Omega, \\ y_0 &= \begin{cases} 0 \text{ on } \Sigma_1 \\ \frac{1}{\beta} \frac{\partial p_0}{\partial \nu} \text{ on } \Sigma_2, p_0 = 0 \text{ on } \Sigma, \\ 0 \text{ on } \Sigma \setminus \Sigma_0. \end{cases} \end{aligned} \quad (3.5)$$

and where  $\{z, q\}$  is given by

$$\begin{aligned} z'' - \Delta z &= 0, \\ q'' - \Delta q &= z, \\ z(0) = z'(0) &= 0, \quad q(T) = q'(T) = 0, \\ z &= \begin{cases} v_1 \text{ on } \Sigma_1 \\ \frac{1}{\beta} \frac{\partial q}{\partial \nu} \text{ on } \Sigma_2, q = 0 \text{ on } \Sigma, \\ 0 \text{ on } \Sigma \setminus \Sigma_0. \end{cases} \end{aligned} \quad (3.6)$$

We set next

$$Lv_1 = \{z'(T; v_1), -z(T; v_1)\} \quad (3.7)$$

which defines

$$L \in \mathcal{L}(L^2(\Sigma_1); H^{-1}(\Omega) \times L^2(\Omega)). \quad (3.8)$$

Then (3.2), using (3.4) and (3.7) can be rewritten as

$$Lv_1 \in \{y^1 - y'_0(T) + \alpha_1 B_{-1}, -y^0 + y_0(T) + \alpha_0 B_0\}. \quad (3.9)$$

We introduce two convex proper functions as follows, firstly

$$F_1(v_1) = \frac{1}{2} \int_{\Sigma_1} v_1^2 d\Sigma \text{ on } L^2(\Sigma_1), \quad (3.10)$$

the second one being given on  $H^{-1}(\Omega) \times L^2(\Omega)$  by

$$F_2(g, h) = \begin{cases} 0 & \text{if } \{g, h\} \in \{y^1 - y'_0(T) + \alpha_1 B_{-1}, -y^0 + y_0(T) + \alpha_0 B_0\}, \\ +\infty & \text{otherwise.} \end{cases} \quad (3.11)$$

With these notations problem (3.1), (3.2) becomes equivalent to

$$\inf_{v_1} F_1(v_1) + F_2(Lv_1) \quad (3.12)$$

provided we prove that the range of  $L$  is dense in  $H^{-1}(\Omega) \times L^2(\Omega)$ , under condition (1.25).  $\square$

We introduce the "adjoint states"  $\varphi, \psi$  as follows. Let  $f^0, f^1$  be given in  $H_0^1(\Omega) \times L^2(\Omega)$ . We define  $\varphi, \psi$  as the unique solution of the system

$$\begin{aligned} \varphi'' - \Delta\varphi &= \psi, \\ \psi'' - \Delta\psi &= 0, \\ \varphi(T) &= f^0, \varphi'(T) = f^1, \psi(0) = \psi'(0) = 0, \\ \varphi &= 0 \text{ on } \Sigma, \psi = \begin{cases} 0 & \text{on } \Sigma_1 \\ \frac{1}{\beta} \frac{\partial\varphi}{\partial\nu} & \text{on } \Sigma_2, \\ 0 & \text{on } \Sigma \setminus \Sigma_0. \end{cases} \end{aligned} \quad (3.13)$$

If we multiply the first (resp. second) equation in (3.13) by  $z$  (resp.  $q$ ), we obtain, after integrations by parts,

$$(z'(T), f^0) - (z(T), f^1) = - \int_{\Sigma_1} \frac{\partial\varphi}{\partial\nu} v_1 d\Sigma. \quad (3.14)$$

Therefore if  $(z'(T), f^0) - (z(T), f^1) = 0 \forall v_1 \in L^2(\Sigma_1)$ , then

$$\frac{\partial\varphi}{\partial\nu} = 0 \text{ on } \Sigma_1. \quad (3.15)$$

Then in case (1.16)  $\psi = 0$  on  $\Sigma$  so that  $\psi \equiv 0$ . Therefore

$$\varphi'' - \Delta\varphi = 0, \varphi = 0 \text{ on } \Sigma \quad (3.16)$$

and satisfies to (3.15). Therefore, according to Holmgren's uniqueness theorem (cf. L Hormander [3], Th. 5.3.3, and for the explicit use of it made here, cf. J L Lions [5], Chapter 1, Section 8) if (1.25) takes place,  $\varphi \equiv 0$ , so that  $f^0 = 0, f^1 = 0$  and (3.3) is proven.  $\square$

We can now apply to (3.12) the duality of Fenchel and T R Rockafellar [9] (cf. also I Ekeland and R Temam [1]). We obtain

$$\inf_{v_1} F_1(v_1) + F_2(Lv_1) = - \inf_{f^0, f^1 \in H_0^1(\Omega) \times L^2(\Omega)} F_1^*(L^*\{f^0, f^1\}) + F_2^*(-f^0, -f^1) \quad (3.17)$$

where  $F_i^*$  is the conjugate function of  $F_i$  and  $L^*$  the adjoint of  $L$ .

By virtue of (3.14)

$$L^*\{f^0, f^1\} = -\frac{\partial \varphi}{\partial v} \text{ on } \Sigma_1. \quad (3.18)$$

We see easily that

$$F_1^* = F_1$$

and

$$F_2^*(f^0, f^1) = (f^0, y^1 - y'_0(T)) + \alpha_1 \|f^0\|_{H_0^1(\Omega)} + (f^1, -y^0 + y_0(T)) + \alpha_0 \|f^1\|_{L^2(\Omega)}.$$

Therefore the (opposite of) right hand side of (3.17) is given by

$$\inf_{f^0, f^1} \left[ \frac{1}{2} \int_{\Sigma_1} \left( \frac{\partial \varphi}{\partial v} \right)^2 d\Sigma_1 - (f^0, y^1 - y'_0(T)) + (f^1, y^0 - y_0(T)) + \alpha_1 \|f^0\|_{H_0^1(\Omega)} + \alpha_0 \|f^1\|_{L^2(\Omega)} \right]. \quad (3.19)$$

This is the dual problem of (3.1), (3.2).  $\square$

We have now two ways to derive the optimality system for the leader control, starting from the primal or from the dual problem. We obtain the following result.

**Theorem 3.1.** *We assume that (1.16) and (1.25) hold true. For  $\{f^0, f^1\}$  in  $H_0^1(\Omega) \times L^2(\Omega)$  we uniquely define  $\{\varphi, \psi, y, p\}$  by:*

$$\varphi'' - \Delta \varphi = \psi, \psi'' - \Delta \psi = 0,$$

$$y'' - \Delta y = 0, p'' - \Delta p = y - y_2, \text{ in } \Omega \times (0, T),$$

$$\varphi(T) = f^0, \varphi'(T) = f^1, \psi(0) = \psi'(0) = 0,$$

$$y(0) = y'(0) = 0, p(T) = p'(T) = 0, \text{ in } \Omega$$

$$\varphi = 0 \text{ on } \Sigma, \psi = \begin{cases} 0 & \text{on } \Sigma_1 \\ \frac{1}{\beta} \frac{\partial \varphi}{\partial v} & \text{on } \Sigma_2, \\ 0 & \text{on } \Sigma_3 \end{cases} \quad (3.20)$$

$$\begin{aligned}
& -\frac{\partial \varphi}{\partial \nu} \text{ on } \Sigma_1 \\
y = & \frac{1}{\beta} \frac{\partial p}{\partial \nu} \text{ on } \Sigma_2, p = 0 \text{ on } \Sigma, \\
& 0 \text{ on } \Sigma \setminus \Sigma_0.
\end{aligned}$$

We uniquely define  $f^0, f^1$  as the solution of the Variational Inequality

$$\begin{aligned}
& (y''(T) - y^1, \hat{f}^0 - f^0) - (y(T) - y^0, \hat{f}^1 - f^1) \\
& + \alpha_1 \|\hat{f}^0\|_{H_0^1(\Omega)} - \alpha_1 \|\hat{f}^0\|_{H_0^1(\Omega)} + \alpha_0 \|\hat{f}^1\|_{L^2(\Omega)} - \alpha_0 \|f^1\|_{L^2(\Omega)} \geq 0 \\
& \forall \hat{f}^0, \hat{f}^1 \in H_0^1(\Omega) \times L^2(\Omega).
\end{aligned} \tag{3.21}$$

Then the optimal leader is given by

$$v_1 = -\frac{\partial \varphi}{\partial \nu} \text{ on } \Sigma_1 \tag{3.22}$$

where  $\varphi$  corresponds to the solution of (3.21).

## References

- [1] Ekeland I and Temam R, *Analyse convexe et problèmes variationnels*, (Dunod, Gauthier Villars), (1974)
- [2] Glowinski R and Lions J L, *Acta Numerica* (1994)
- [3] Hörmander L, *Linear partial differential operators*, (Springer-Verlag) (1976)
- [4] Lions J L, *Contrôle des systèmes distribués singuliers*, (Paris, Gauthier-Villars) (1983)
- [5] Lions J L, *Contrôlabilité exacte, perturbations et stabilisation de systèmes distribués*, (Tome 1 et Tome 2, Masson, RMA) (1988)
- [6] Lions J L, *Some methods in the mathematical analysis of systems and their control*, (Science Press, Beijing and Gordon and Breach, Sc. Pub., New-York) (1981)
- [7] Lions J L, Some remarks on Stackleberg optimization, in *Mathematical Models and Methods in Applied Sciences*. (1993)
- [8] Lions J L, *Contrôle optimal des systèmes gouvernés par des équations aux dérivées partielles*, (Dunod, Gauthier-Villars, Paris) (1968)
- [9] Rockafellar T R, Duality and stability in extremum problems involving convex functions, *Pac. J. Math.* **21** (1967) 167-187
- [10] Stackleberg Von, *Marktform und Gleichgewicht* (Vienna, Julius Springer) (1934)

## Bertini theorems for ideals linked to a given ideal

TRIVEDI VIJAYLAXMI

School of Mathematics, Tata Institute of Fundamental Research, Homi Bhabha Road,  
 Bombay 400 005, India

MS received 30 January 1993; revised 21 March 1993

**Abstract.** We prove a generalization of Flenner's local Bertini theorem for complete intersections. More generally, we study properties of the 'general' ideal linked to a given ideal. Our results imply the following. Let  $R$  be a regular local Nagata ring containing an infinite perfect field  $k$ , and  $I \subset R$  is an equidimensional radical ideal of height  $r$ , such that  $R/I$  is Cohen-Macaulay and a local complete intersection in codimension 1. Then for the 'general' linked ideal  $J_\alpha$ ,  $R/J_\alpha$  is normal and Cohen-Macaulay.

The proofs involve a combination of the method of basic elements, applied to suitable blow ups.

**Keywords.** Linkage; local Bertini theorem; Nash blow up;  $r$ -fold basic elements.

### 1. Introduction

Let  $R$  be a Noetherian local ring, and  $I \subset R$  an ideal of height  $r$ . This paper is concerned with questions of the following type: if  $J$  is a "sufficiently general" ideal in  $R$  which is geometrically linked to  $I$  (see §6 for definition), then what "good properties" does the ring  $R/J$  inherit from  $R$ , or from  $R/I$ ?

Part of the problem is to make precise the notion of "sufficiently general". We will work with the following natural notion, motivated by the standard one in algebraic geometry. We assume that our local ring  $R$  contains an infinite field  $k$ . Suppose  $g_1, \dots, g_N$  is a set of generators for  $I$ . Then we can consider the set of  $r$ -tuples of  $k$ -linear combinations  $\{\sum_j \alpha_{ij} g_j \mid 1 \leq i \leq r\}$  as the ( $k$ -valued) points of the affine space  $A^{rN}$ . For any such  $r$ -tuple  $\alpha \in A^{rN}$ , let  $\mathfrak{F}_\alpha = \sum_i R(\sum_j \alpha_{ij} g_j)$  be the ideal generated by the  $r$ -tuple. Define  $J_\alpha = (\mathfrak{F}_\alpha : I)$ . We will say that some property holds *for the general*  $J_\alpha$  (or just *for general*  $\alpha$ ) if the property is valid for all points  $\alpha$  in a non-empty Zariski open subset of  $A^{rN}$ .

For  $r=1$ , this notion of "general" is used by Flenner in formulating some of his local Bertini theorems. For example our Theorem 2 (see section 4) is a generalisation of the following theorem of Flenner ([F], Satz 4.6), apart from the fact that Flenner's proof works for any set of generators of  $I$ .

**Theorem (Flenner).** *Let  $(A, m)$  be a Noetherian local  $k$ -algebra, where  $k$  is a field of characteristic 0. Let  $I \subset A$  be an ideal, with a given set of generators  $x_1, \dots, x_N$ . Then for general  $\alpha \in k^N$ , the quotient ring  $R/x_\alpha R$  is regular on  $D(I) \cap \text{reg}(R)$ . Here  $x_\alpha = \sum_j \alpha_j x_j$ .*

We show, by an example in §4, that if  $k$  has characteristic  $p > 0$ , then Theorem 2 above does not hold for an arbitrary set of generators of  $I$ .

$T \subset \mathbb{A}^M \times \mathbb{A}^N$  is a subset such that for some non-empty (Zariski) open subset  $U \subset \mathbb{A}^M$ , and each  $x \in U$ ,  $T \cap \{x\} \times \mathbb{A}^N$  contains a non-empty open subset of  $\{x\} \times \mathbb{A}^N$ , then it need not be true that  $T$  contains a non-empty open subset of  $\mathbb{A}^M \times \mathbb{A}^N$ . In fact, Flenner's argument seems to use some special features of the case  $r = 1$  which do not generalise to the case  $r > 1$ .

There is a related notion of the "generic" ideal  $J$ , corresponding to adjoining  $rN$  indeterminates  $X_{ij}$  to  $R$ , and considering the corresponding  $r$ -tuple of linear combinations  $\{\sum_j X_{ij} g_j \mid 1 \leq i \leq r\}$ . This corresponds to choosing the *generic point*, in the sense of algebraic geometry, of the affine space  $\mathbb{A}^{rN}$  considered above. Bertini theorems for the generic linked ideal have been obtained earlier by Flenner (*loc. cit.*), and Huneke and Ulrich [HU].

We will mainly be concerned with properties of the following type: the property will be local on  $\text{Spec } R$ , and the subset  $\{(\alpha, \mathfrak{p}) \mid \mathbb{A}^{rN} \times \text{Spec } R\}$  of pairs where  $R/J_\alpha$  does not have the property at  $\mathfrak{p}$ , is a Zariski closed (or more generally, locally closed) subscheme of  $\mathbb{A}^{rN} \times \text{Spec } R$ . For example, we may ask if  $R/J_\alpha$  contains the point  $\mathfrak{p}$  and is regular at  $\mathfrak{p}$ ; results in this direction are Bertini theorems.

For local rings which are localisations of finitely generated algebras over a field, standard theorems (essentially, Chevalley's theorem on constructible sets) show that the two notions (of "general" as defined above, and "generic") are equivalent, for such properties. However, this is by no means clear for more general local rings  $R$ , like complete local rings.

The main technical innovation in this paper is a way of showing that certain properties are valid for the *general* ideal  $J_\alpha$ , as follows. Using appropriate blow ups (in particular, what we call the *Nash blow up*), we reduce the proof of the desired property for the general  $J_\alpha$  to proving certain other properties for general  $r$ -tuples of sections of locally free sheaves on projective varieties over a field. We are then in a position to use results from algebraic geometry. This strategy may have applications in other problems where one wants to consider the "general" rather than the "generic".

The main results proved in the paper are Theorems 2 and 3, which are, roughly speaking, the Bertini theorems "on  $D(I)$ " and "on  $V(I)$ " (here  $V(I) = \text{Spec } R/I \subset \text{Spec } R$ , and  $D(I) = \text{Spec } R - V(I)$ ). Since the precise statements are a bit technical, we instead state here a consequence.

**Theorem.** *Let  $R$  be a local Cohen–Macaulay Nagata ring containing an infinite perfect field  $k$ ,  $I \subset R$  an ideal of height  $r$ . Then  $I$  has a set of generators  $\{g_1, \dots, g_N\}$  such that if  $f_i = \sum_j \alpha_{ij} g_j$ ,  $1 \leq i \leq r$ , and  $\mathfrak{F}_\alpha = (f_1, \dots, f_r)$ ,  $J_\alpha = (\mathfrak{F}_\alpha : I)$  are as defined above (so that  $J_\alpha$  is the "general" ideal linked to  $I$ ), then*

- (i)  $f_1, \dots, f_r$  form a regular sequence
- (ii) the ring  $R/J_\alpha$  is regular at all primes of  $D(I) \cap \text{reg}(R)$
- (iii) if the ring  $R/I$  is reduced equidimensional and is a local complete intersection in codimension 1, then the ring  $(R/J_\alpha)_\mathfrak{p}$  is regular for all  $\mathfrak{p} \in \{\text{reg}(R) \cap \text{codim.1 ideals of } R/J_\alpha\}$ .

We state a corollary.



## COROLLARY

Let  $R$  be a regular local Nagata ring containing an infinite perfect field  $k$ ,  $I \subset R$  an equidimensional radical ideal of height  $r$ , such that

- (i)  $R/I$  is Cohen-Macaulay, and
- (ii)  $R/I$  is a local complete intersection in codimension 1.

Then for the "general" ideal  $J_\alpha$ ,  $R/J_\alpha$  is normal and Cohen-Macaulay.

The paper is organised as follows. In § 2, we collect together some technical lemmas on blow ups and determinantal schemes, and a regularity criterion. In § 3, we prove that the general ideal  $\mathfrak{F}_\alpha$  has height  $r$ , by reducing this to Noether normalisation. In § 4, we prove the regularity assertion of  $J_\alpha$  on  $D(I) \cap \text{reg}(R)$ . In § 5, we prove a Bertini theorem for modules which is the main step in the proof of the regularity assertion for  $R/J_\alpha$  along  $V(I)$ . In the last section, we deduce consequences of the earlier results, which include the above theorem.

## 2. Some lemmas

In this section, we prove several lemmas which are needed later. Most of these are folklore, and the proofs use standard arguments. However, I was unable to find suitable references where they are proved in the desired form, and so have given the proofs.

*Lemma 1.* Let  $(R, \mathfrak{m})$  be a local domain. Let  $M$  be a torsion free  $R$ -module of rank  $r$ . If there exists a surjective map of  $R$ -modules  $\wedge^r M \rightarrow R$ , then  $M \cong R^r$ .

*Proof.* We prove the result by induction on  $r$ .

Let  $r = 1$ . The existence of some surjective homomorphism  $M \rightarrow R$  implies that  $M = N \oplus R$ , for some  $R$ -module  $N$ . Let  $Q(R)$  denote the quotient field of  $R$ . Now  $Q(R) \cong Q(R) \otimes M = (Q(R) \otimes N) \oplus Q(R)$ , which implies  $Q(R) \otimes N = 0$ . But  $N$  is a submodule of the torsion free module  $M$ , therefore is zero. Hence  $M = R$ .

Assume the lemma for  $r - 1$ .

*Claim.* If  $M_1$  and  $M_2$  are two  $R$ -modules such that  $M_1 \otimes M_2$  has a nonzero free direct summand then  $M_1$  also has one.

We assume the claim for the moment.

Consider the canonical  $R$ -linear maps

$$M \otimes \wedge^{r-1} M \rightarrow \wedge^r M \rightarrow R.$$

By the claim,  $M \cong M_1 \oplus R$ , where  $M_1$  is a torsion free module of rank  $r - 1$ . Now  $\wedge^r M \cong \wedge^r(M_1 \oplus R) \cong \wedge^{r-1} M_1$ . Hence by induction,  $M_1$  is free; therefore so is  $M$ .

*Proof of the claim.* Since the module  $M_1 \otimes M_2$  has a free direct summand, there exists a surjective map  $\phi: M_1 \otimes M_2 \rightarrow R$ . Consider a surjective map  $\psi: R^{\oplus n} \rightarrow M_2$ , where  $n$  need not be finite. Therefore  $\phi \circ (Id \otimes \psi): M_1^{\oplus n} \rightarrow R$  is surjective, which implies that this map restricted to at least one summand  $M_1$  is surjective. Hence the claim.  $\square$

**Lemma 2.** Let  $R$  be a Noetherian integral domain and let  $M$  be a finitely generated  $R$ -module of rank  $t$ . Let  $S = \text{Sym}_R(\wedge^t M)/(R\text{-torsion})$  and  $X = \text{Proj}(S)$ . Then  $S$  is a domain and the sheaf  $\tilde{M}'$ , associated to  $M' := (M \otimes S)/(S\text{-torsion})$  on  $X$ , is a locally free  $\mathcal{O}_X$ -module of rank  $t$ .

*Proof.* Let  $K$  denote the quotient field of  $R$ . Then the canonical  $R$ -linear localisation map  $S = \text{Sym}_R(\wedge^t M)/(R\text{-torsion}) \rightarrow \text{Sym}_K(\wedge^t M \otimes_R K) = K[T]$  is injective, as the  $R$ -module  $S$  is torsion free by definition. This implies that  $S$  is a domain.

Consider the canonical map  $\wedge^t M \otimes_R S \rightarrow S(1)$ . This induces a map  $\wedge^t_S(M \otimes_R S)/(R\text{-torsion}) \rightarrow S(1)$  of graded  $S$ -modules, as  $S$  is integral. Which induces a map of  $\mathcal{O}_X$ -modules  $(\wedge^t_S \tilde{M}') \rightarrow \tilde{S}(1)$ . This is a surjective map, and may be considered as the map  $\wedge^t_S(\tilde{M}') \rightarrow \tilde{S}(1)$ ; therefore, by the above lemma,  $\tilde{M}'$  is a locally free  $\mathcal{O}_X$ -module of rank  $t$ .  $\square$

**Remark (1).** With the notations of the previous lemma let  $\pi: \text{Proj}(S) \rightarrow \text{Spec}(R)$  be the canonical map. Then the open subset  $U_R := \{\mathfrak{p} \in \text{Spec}(R) | M_{\mathfrak{p}} \text{ is a free } R_{\mathfrak{p}}\text{-module}\}$  can be covered by sets  $\{U_i\}_i$ , where each  $U_i = \text{Spec}(R_i)$  is an affine open set such that the module  $M \otimes R_i$  is a free  $R_i$ -module. Then the map  $\pi|_{U_i}: X_{U_i} = \text{Proj } R_{U_i}[T] \rightarrow \text{Spec } R_{U_i}$  is an isomorphism; which implies that the map  $\pi: \pi^{-1}(U) \rightarrow U$  is an isomorphism. In general, if  $Z$  is a Noetherian integral scheme and  $\mathcal{F}$  is a rank  $t$  coherent sheaf of  $\mathcal{O}_Z$  modules, let  $X = \text{Proj}(\text{Sym}_{\mathcal{O}_Z}(\wedge^t \mathcal{F})/(R\text{-torsion}))$  and let  $W$  be the open subset of  $Z$  consisting of points where  $\mathcal{F}$  is free. Then for the map  $\pi: X \rightarrow Z$ , the restriction map  $\pi: \pi^{-1}(W) \rightarrow W$  is an isomorphism. We call  $\pi: X \rightarrow Z$  the *Nash blow up* of  $Z$  for the sheaf  $\mathcal{F}$ . From the above lemma, it follows that  $\pi^* \mathcal{F}/(R\text{-torsion})$  is a locally free  $\mathcal{O}_X$ -module of rank  $t$ .

## DEFINITION

Let  $R$  be a ring and  $M$  be a finitely generated  $R$ -module. Then a submodule  $N$  of  $M$  is said to be  *$r$ -fold basic* at  $\mathfrak{p} \in \text{Spec}(R)$  if  $\mu_{\mathfrak{p}}(M) - \mu_{\mathfrak{p}}(M/N) \geq r$ , i.e.,  $N_{\mathfrak{p}}$  contains at least  $r$  minimal generators of  $M_{\mathfrak{p}}$ .

In particular, 'a subset  $\{a_1, a_2, \dots, a_r\}$  of  $M$  generates an  $r$ -fold basic sub-module of  $M$  at  $\mathfrak{p}$ ' means it is a part of minimal set of generators of  $M_{\mathfrak{p}}$  equivalently the image of  $\{a_1, a_2, \dots, a_r\}$  in  $(M/\mathfrak{p}M)_{\mathfrak{p}}$  generates an  $r$  dimensional  $(R/\mathfrak{p})_{\mathfrak{p}}$ -subspace.

These definitions can be extended to coherent sheaves on schemes in an obvious way.

**Lemma 3.** With the notations of the previous remark, let  $Y = \pi^{-1}(\mathfrak{m})$  be the reduced closed subscheme of  $X := \text{Proj } S$  with the natural closed immersion  $i: Y \rightarrow X$ ; let  $\mathcal{F}$  be a locally free sheaf of  $\mathcal{O}_X$ -modules and  $\mathcal{E} := i^* \mathcal{F}$ . If a given set of sections  $\{h_1, \dots, h_r\}$  of  $\mathcal{F}$  generates an  $r$ -fold basic  $\mathcal{O}_Y$ -submodule of  $\mathcal{E}$  then it generates an  $r$ -fold basic submodule of  $\mathcal{F}$  also.

*Proof.* For  $\mathfrak{q} \in X$ , consider some maximal homogenous prime ideal  $\mathfrak{q}'$  containing  $\mathfrak{q}$  in  $X = \text{Proj } S$ ; then  $\mathfrak{q}'$  is a closed point of  $X$ . Since  $\pi$  is a closed map,  $\mathfrak{q}' \mapsto \mathfrak{m} \in \text{Spec}(R)$ , hence  $\mathfrak{q}' \in \pi^{-1}(\mathfrak{m}) = Y$ . If  $\{h_1, \dots, h_r\}$  generates an  $r$ -fold basic  $\mathcal{O}_Y$ -submodule of  $\mathcal{E}$  at  $\mathfrak{q}'$ , then  $\{h_1, \dots, h_r\}$  is an  $r$ -fold basic  $\mathcal{O}_X$ -submodule of  $\mathcal{F}$  at  $\mathfrak{q}'$ , and hence also at  $\mathfrak{q}$  as  $\mathcal{F}$  is locally free.  $\square$

**Lemma 4.** *With the notations of the lemma 2, if  $h_1, \dots, h_r$  are elements of  $M$  such that they generate an  $r$ -fold basic  $\mathcal{O}_X$ -submodule of  $\tilde{M}'$  then they are  $r$ -fold basic elements of  $M$ .*

*Proof.* As  $\pi$  is a closed map and the generic point of  $X$  maps to the generic point of  $R$ , the map  $\pi: X \rightarrow \text{Spec } R$  is surjective. Therefore, for given  $\mathbf{p} \in \text{Spec } R$ , one can choose  $\mathbf{q} \in X$  such that  $\pi(\mathbf{q}) = \mathbf{p}$ . Hence there is the canonical inclusion of fields

$$k(\mathbf{p}) := R_{\mathbf{p}}/\mathbf{p}R_{\mathbf{p}} \rightarrow k(\mathbf{q}) := \mathcal{O}_{X,\mathbf{q}}/\mathbf{q}\mathcal{O}_{X,\mathbf{q}}$$

and the canonical  $k(\mathbf{p})$ -linear map, say

$$\psi: M \otimes_R k(\mathbf{p}) \rightarrow \tilde{M}' \otimes_{\mathcal{O}_X} k(\mathbf{q})$$

where  $M \otimes_R k(\mathbf{p})$  is a  $k(\mathbf{p})$ -vector space and  $\tilde{M}' \otimes_{\mathcal{O}_X} k(\mathbf{q})$  is a  $k(\mathbf{q})$ -vector space. Since  $h_1, \dots, h_r$  are  $r$ -fold basic  $\mathcal{O}_X$ -submodule of  $\tilde{M}'$  at  $\mathbf{q}$ , the elements  $\psi(h_1), \dots, \psi(h_r)$  are linearly independent elements of  $\tilde{M}' \otimes_{\mathcal{O}_X} k(\mathbf{q})$ . This implies that  $h_1, \dots, h_r$  are linearly independent elements of  $M \otimes k(\mathbf{p})$ . This assertion is the same as saying that  $h_1, \dots, h_r$  generates an  $r$ -fold basic submodule of  $M$  at  $\mathbf{p}$ . Hence the lemma.  $\square$

**Lemma 5.** *Let  $R = k[x_1, \dots, x_r, y_{11}, \dots, y_{rd}]$  be a polynomial ring in  $r(d+1)$  variables, and let  $I$  be the ideal generated by the  $r \times r$  minors of the matrix*

$$\begin{bmatrix} x_1 & 0 & y_{11} & \cdots & y_{1d} \\ & \ddots & \vdots & & \vdots \\ 0 & x_r & y_{r1} & \cdots & y_{rd} \end{bmatrix}.$$

*Then  $\dim R/I = (r-1)(d+1)$ .*

*Proof.* Let  $\mathbf{A}^r \times \mathbf{A}^{rd} = \text{Spec}(R)$  and let  $W = V(I)$  denote the subvariety of  $\mathbf{A}^r \times \mathbf{A}^{rd}$  given by the ideal  $I$ . Consider the projection map  $p: \mathbf{A}^r \times \mathbf{A}^{rd} \rightarrow \mathbf{A}^r$  defined as  $(x_1, \dots, x_r, y_{11}, \dots, y_{rd}) \rightarrow (x_1, \dots, x_r)$ . Then  $p|_W: W \rightarrow D \subset \mathbf{A}^r$  is a surjective map, where  $D = \{(x_1, \dots, x_r) \in \mathbf{A}^r \mid x_1 \cdots x_r = 0\}$ . Let  $T$  run over the proper subsets of  $\{1, \dots, r\}$ ; then the set  $D$  is covered by finite number of locally closed sets of the type

$$D_T = \{(x_1, \dots, x_r) \in D \mid x_j \neq 0 \text{ for } j \in T \text{ and the other } x_i\text{'s are zero}\}.$$

It is easy to see that  $\dim(D_T) = \text{card}(T)$ .

Consider a typical such set  $D_{T_0} = \{(x_1, \dots, x_r) \in D \mid x_1 \cdots x_k \neq 0 \text{ and } x_{k+1} = \dots = x_r = 0\}$ . Let  $\mathbf{x} = (x_1, \dots, x_r)$ . If  $W_{\mathbf{x}} \subset \mathbf{A}^{rd}$  denotes the fibre of  $W$  over the point  $\mathbf{x}$ , then for  $\mathbf{x} \in D_{T_0}$ , the ideal of this fibre in  $k[y_{ij}]$  is

$$I(W_{\mathbf{x}}) = \text{ideal generated by } (r-k) \text{ minors of } \begin{bmatrix} y_{(k+1)1} & \cdots & y_{(k+1)d} \\ \vdots & & \vdots \\ y_{r1} & \cdots & y_{rd} \end{bmatrix}.$$

By [EH], corollary (4)

$$\begin{aligned} \text{ht } I(W_{\mathbf{x}}) &= (r-k - (r-k-1))(d - (r-k-1)) \\ &= d - r + k + 1 \quad (\text{for } 0 \leq k \leq r-1) \end{aligned}$$

Hence for every  $x \in D_{T_0}$ ,

$$\begin{aligned}\dim W_x &= rd - d + r - k - 1 \\ &= (r-1)(d+1) - k\end{aligned}$$

Thus for  $x \in D_T$ ,  $\dim W_x = (r-1)(d+1) - \text{card}(T)$ . It is easy to see that  $\dim W = \sup_T \{\dim p^{-1}(D_T)\}$ , where  $p: W \rightarrow D$  is the canonical map. Since for the surjective map  $p: p^{-1}(D_T) \rightarrow D_T$ , the dimension of each fibre is  $(r-1)(d+1) - \text{card}(T)$ , we have (see [Sh], Chapter I, §6, Theorem 7, for example)

$$\begin{aligned}\dim p^{-1}(D_T) &= (r-1)(d+1) - \text{card}(T) + \text{card}(T) \\ &= (r-1)(d+1).\end{aligned}$$

Therefore  $\dim W = (r-1)(d+1)$ . □

Let  $R$  be a Noetherian ring,  $M$  a finite  $R$ -module. For any  $R$ -algebra  $S$ , and any  $S$ -linear map  $\psi: S \rightarrow M \otimes S$ , let  $I(\psi)$  be the ideal  $\psi^*((M \otimes S)^*) \subset S$ , where  $N^*$  denotes the dual  $\text{Hom}_S(N, S)$ , for any  $S$ -module  $N$ . Let

$$T_S^k = \{\mathfrak{p} \in \text{Spec}(S) \mid M \otimes S_{\mathfrak{p}} \text{ is free and } S_{\mathfrak{p}} \text{ is regular of } \dim \leq k\}.$$

*Lemma 6.* Let  $(R, \mathfrak{m})$  be a local Nagata ring and let  $M$  be a finitely generated  $R$ -module. Let  $\psi: R \rightarrow M$  be an  $R$ -linear map.

- (i) Let  $\text{Spec}(R) = \cup_i \text{Spec}(R_i)$  be the decomposition into irreducible components,  $M_i = M \otimes R_i$  and  $\psi_i = \psi \otimes M_i$ , and suppose  $R_i/I(\psi_i)$  is regular on  $T_{R_i}^k$  for each  $i$ . Then  $R/I(\psi)$  is regular on  $T_R^k$ .
- (ii) Let  $\hat{R}$  be the completion of  $R$ , and  $\hat{\psi} = \psi \otimes \text{Id}: \hat{R} \rightarrow \hat{M} = M \otimes \hat{R}$ , and suppose  $\hat{R}/I(\hat{\psi})$  is regular on  $T_{\hat{R}}^k$ . Then  $R/I(\psi)$  is regular on  $T_R^k$ .

*Proof.* i) Let  $\eta_i: \text{Spec}(R_i) \hookrightarrow \text{Spec}(R)$  be the canonical map. The set  $T_{R_i}^k \supseteq \text{Spec}(R_i) \cap T_R^k$ , and for the induced map  $\psi_i: R_i \rightarrow M \otimes R_i$  the ideal  $I(\psi_i) = \psi_i^*((M \otimes R_i)^*) = \psi^*(M^*) \otimes R_i$ . Since  $T_R^k \subseteq \cup T_{R_i}^k$ , for any  $\mathfrak{p} \in T_R^k$  there exists some  $R_i$  for which  $\mathfrak{p} \in T_{R_i}^k$ . Then we have  $(R_i)_{\mathfrak{p}} = R_{\mathfrak{p}}$  and  $I(\psi_i)_{\mathfrak{p}} = I(\psi)_{\mathfrak{p}}$ .

ii) Since the map  $R \rightarrow \hat{R}$  is a faithfully flat extension and  $R$  is a Nagata ring, we have  $\eta(\text{Reg}(\hat{R})) \supseteq \text{Reg}(R)$ , where  $\eta: \text{Spec}(\hat{R}) \rightarrow \text{Spec}(R)$  is the canonical map. Therefore  $\eta(T_{\hat{R}}^k) \supseteq T_R^k$ . Now, for the canonical map  $\psi \otimes \hat{R}: \hat{R} \rightarrow M \otimes \hat{R}$ , we have  $I(\hat{\psi}) = (\psi \otimes \hat{R})^*((M \otimes \hat{R})^*) = \psi^*(M^*) \otimes \hat{R} = I(\psi) \otimes \hat{R}$ . If  $\hat{R}/I(\hat{\psi})$  is regular at  $T_{\hat{R}}^k$  then  $R/I(\hat{R})$  is regular on  $T_R^k$  as  $R/I(\psi) \rightarrow \hat{R}/I(\hat{\psi})$  is a faithfully flat extension. □

The next lemma is a slight extension of a lemma of Flenner (see [F], (2.2)).

*Lemma 7.* Let  $(R, \mathfrak{m})$  be a local ring and let  $x_1, \dots, x_r \in \mathfrak{m}$ . Let  $d: R \rightarrow M$  be a derivation, where  $M$  is a finitely generated  $R$ -module, such that  $\langle d(x_1), \dots, d(x_r) \rangle$  is an  $r$ -fold basic submodule of  $M$  at  $\mathfrak{m}$ . Then  $\{x_1, \dots, x_r\}$  is a part of a minimal set of generators for  $\mathfrak{m}$ . In particular, if  $R$  is regular, so is  $R/(x_1, \dots, x_r)$ .

*Proof.* Since  $d(x_1)$  is a basic element of  $M$ , we have  $x_1 \notin \mathfrak{m}^2$ , for if  $x_1 = \sum r_i s_i$  where  $r_i, s_i \in \mathfrak{m}$  then  $d(x) = \sum s_i d(r_i) + \sum r_i d(s_i) \in \mathfrak{m}M$ . Therefore the ring  $R/(x_1)$  is regular. Let

$M_1 = M/(R(dx_1) + (x_1)M)$  and let  $d_1: R/(x_1) \rightarrow M_1$  be the canonical derivation. Since  $M_1/\mathfrak{m}M_1 = M/(\mathfrak{m}M + R(dx_1))$ , the submodule  $\langle d_1(\bar{x}_2), \dots, d_1(\bar{x}_r) \rangle$  of  $M_1$  is  $(r-1)$ -fold basic, where  $\bar{x}_i$  is the image of  $x_i$  in  $R/(x_1)$ . Now, by induction on  $r$ , we conclude that  $(R/(x_1))/(\bar{x}_2, \dots, \bar{x}_r) = R/(x_1, \dots, x_r)$  is regular.  $\square$

**Lemma 8.** Let  $R$  be a Noetherian ring containing an infinite field  $k$ . Let  $F$  be a finite set in  $\text{Spec } R$ . Let  $M$  be an  $R$ -module generated by  $g_1, \dots, g_N$  and let

$$\phi_i = \sum_j \alpha_{ij} g_j \in M, \quad 1 \leq i \leq r,$$

where  $(\alpha_{ij}) \in k^{rN}$ . Let  $M'_\alpha = \langle \phi_1, \dots, \phi_r \rangle$  be the corresponding submodule of  $M$ , and let  $\Phi = \phi_1 \wedge \dots \wedge \phi_r \in \wedge^r M$ .

Assume that for each  $\mathfrak{p} \in F$ , the  $R_\mathfrak{p}$ -module  $M_\mathfrak{p}$  is free of rank  $r$ , so that  $\wedge^r M_\mathfrak{p} \cong R_\mathfrak{p}$ . Let  $T_{\alpha, \mathfrak{p}}$  be the ideal in  $R_\mathfrak{p}$  such that the  $R_\mathfrak{p}$ -submodule  $R_\mathfrak{p}\Phi$  equals  $T_{\alpha, \mathfrak{p}} \wedge^r M_\mathfrak{p}$ .

Then for all  $(\alpha_{ij})$  in a non-empty Zariski open subset of  $k^{rN}$ , the ideal  $T_{\alpha, \mathfrak{p}} = R_\mathfrak{p}$  and hence  $(M'_\alpha)_\mathfrak{p} = M_\mathfrak{p}$ , for all  $\mathfrak{p} \in F$ .

*Proof.* For  $\mathfrak{p} \in F$ , the module  $M_\mathfrak{p}$  is freely generated by  $r$  elements from the chosen set of generators  $\{g_1, \dots, g_N\}$ , say by  $g_1, \dots, g_r$ . Let  $g_k = a_{1k}g_1 + \dots + a_{rk}g_r$ , where  $k \geq r+1$  and  $a_{jk} \in R_\mathfrak{p}$ . Then one computes that  $T_{\alpha, \mathfrak{p}} := \det([\alpha_{ij} + \sum_{k \geq r+1} \alpha_{ik} a_{jk}]_{r \times r})$  generates  $T_{\alpha, \mathfrak{p}}$ .

The ideal  $\langle T_{\alpha, \mathfrak{p}} \rangle = R_\mathfrak{p}$

$\Leftrightarrow \text{Im}(T_{\alpha, \mathfrak{p}}) \neq 0$  in  $(R/\mathfrak{p})_\mathfrak{p}$

$\Leftrightarrow$  the polynomial  $T_{X, \mathfrak{p}} := \det([X_{ij} + \sum_k \bar{a}_{jk} X_{ik}]_{r \times r}) \in (R/\mathfrak{p})_\mathfrak{p}[\{X_{ij}\}_{r \times N}]$  does not vanish at  $(\alpha_{ij})$ , where  $\bar{a}_{jk}$  denotes the image of  $a_{jk}$  in  $(R/\mathfrak{p})_\mathfrak{p}$  and  $X_{ij}$ 's are indeterminates.

But the zero set of  $T_{X, \mathfrak{p}}$  is a proper subset of  $((R/\mathfrak{p})_\mathfrak{p})^{rN}$ , as the point with coordinates  $X_{ij} = \delta_{ij}$  (Kronecker delta) does not lie in the zero set. Since  $k$  is an infinite field contained in  $(R/\mathfrak{p})_\mathfrak{p}$ , the zero set of  $T_{X, \mathfrak{p}}$  in  $k^{rN}$  is a proper Zariski closed set.

As  $k$  is infinite, a finite intersection of nonempty open subsets of  $k^{rN}$  is a nonempty open set. Therefore for such a given finite set  $F$ , we have  $\langle T_{\alpha, \mathfrak{p}} \rangle = R_\mathfrak{p}$  and hence  $M'_\mathfrak{p} = M_\mathfrak{p}$ , for general  $(\alpha) \in k^{rN}$  and all  $\mathfrak{p} \in F$ .  $\square$

**Lemma 9.** Let  $(R, \mathfrak{m})$  be a Noetherian local ring and let  $[f_{ij}]_{k \times k}$  be a matrix with entries in  $R$ . If  $\det([f_{ij}]) \notin \mathfrak{m}^2$  then at least one of the minors of size  $(k-1)$  of  $[f_{ij}]$  is invertible in  $R$ .

*Proof.* Suppose none of the minors of size  $(k-1)$  of  $[f_{ij}]$  is invertible. Then the matrix  $[\bar{f}_{ij}]$  is of rank  $r \leq k-2$ , where  $\bar{f}_{ij}$  denotes the image of  $f_{ij}$  in  $R/\mathfrak{m}$ . Therefore there exist invertible matrices  $[\bar{g}_{ij}]_{k \times k}$  and  $[\bar{h}_{ij}]_{k \times k}$  over  $R/\mathfrak{m}$  such that

$$[\bar{g}_{ij}] \cdot [\bar{f}_{ij}] \cdot [\bar{h}_{ij}] = \begin{bmatrix} Id_r & 0 \\ 0 & 0 \end{bmatrix},$$

where  $Id_r$  is the  $r \times r$  identity matrix. If elements  $g_{ij}, h_{ij}$  are some lifts of the elements  $\bar{g}_{ij}, \bar{h}_{ij}$  to  $R$  respectively, then  $\det([g_{ij}][f_{ij}][h_{ij}]) \in \mathfrak{m}^2$  as  $r \leq k-2$ . Since  $\det([g_{ij}]), \det([h_{ij}])$  are invertible in  $R$ , we have  $\det([f_{ij}]) \in \mathfrak{m}^2$ , which contradicts the hypothesis.  $\square$

Let  $X$  be a  $k$ -scheme, and let  $\mathcal{E}$  be a locally free sheaf of rank  $r$  generated by the global sections  $\{g_1, \dots, g_N\}$ . Let  $A^{rN} = \text{Spec } k[Y_{11}, \dots, Y_{1N}, \dots, Y_{r1}, \dots, Y_{rN}]$  and let

$$v_i := g_1 Y_{i1} + g_2 Y_{i2} + \dots + g_N Y_{iN} \text{ for } 1 \leq i \leq r. \quad (1)$$

Then we define  $C_0(X) \subset A^{rN} \times X$  to be the zero scheme of the section  $v_1 \wedge \dots \wedge v_r$ . It is easy to see that the ideal sheaf of  $C_0(X)$  is given by

$$\mathcal{I}_{C_0(X)} = \text{image}(v_1 \wedge \dots \wedge v_r)^*. \quad (2)$$

Tensoring the universal  $k$ -derivation map  $\mathcal{O}_X \rightarrow \Omega_{X/k}^1$  with  $\mathcal{O}_{A^{rN}}$ , where  $\Omega_{X/k}^1$  denotes the sheaf of Kähler differentials (see [Ha], II, Sec. 8), we get an  $\mathcal{O}_{A^{rN}}$ -derivation

$$p_2^*(\mathcal{O}_X) \xrightarrow{d} p_2^*(\Omega_{X/k}^1).$$

Now the canonical composite map

$$\mathcal{I}_{C_0(X)} \hookrightarrow p_2^*(\mathcal{O}_X) \xrightarrow{d} p_2^*(\Omega_{X/k}^1), \quad (3)$$

induces an  $\mathcal{O}_{A^{rN} \times X}$  linear map

$$\mathcal{I}_{C_0(X)} / \mathcal{I}_{C_0(X)}^2 \xrightarrow{\bar{d}} p_2^*(\Omega_{X/k}^1) \otimes \mathcal{O}_{C_0(X)}. \quad (4)$$

**Lemma 10.** Let  $Gr$  denote  $Gr_k(r, N)$ , the Grassmannian of  $r$ -dimensional quotients of  $k^N$ . Let  $C_0(Gr)$  be the closed subscheme of  $A^{rN} \times Gr$  as defined above corresponding to the universal quotient bundle  $\mathcal{Q}$ . Let

$$\bar{d}: \mathcal{I}_{C_0(Gr)} / \mathcal{I}_{C_0(Gr)}^2 \rightarrow p_2^*(\Omega_{Gr/k}^1) \otimes_k \mathcal{O}_{C_0(Gr)}$$

be the map defined above. Let  $g \in Gr$ , and  $L$  be a field extension of  $k(g)$ ; let  $B := \Gamma(\mathcal{O}_{C_0(Gr)} \otimes_{\mathcal{O}_g} L)$ . Then

- (i) for a suitable choice of coordinates on  $A^{rN}$ , the ideal of  $C_0(Gr)_g \subset A_{k(g)}^{rN}$  is principal, generated by  $\det(t_{ij})$ , where  $t_{ij}$ ,  $1 \leq i, j \leq r$  are part of the system of coordinates, and
- (ii) the image of

$$\bar{d} \otimes 1: \mathcal{I}_{C_0(Gr)} / \mathcal{I}_{C_0(Gr)}^2 \otimes_{\mathcal{O}_{Gr}} k(g) \rightarrow p_2^*(\Omega_{Gr/k}^1) \otimes_k \mathcal{O}_{C_0(Gr)} \otimes_{\mathcal{O}_g} k(g)$$

is not in the kernel of any nonzero  $B$ -linear map  $B^{r(N-r)} \rightarrow B^m$ , where  $B^m$  denotes any free  $B$ -module and the free  $B$ -module  $p_2^*(\Omega_{Gr/k}^1) \otimes_k \mathcal{O}_{C_0(Gr)} \otimes_{\mathcal{O}_g} k(g) \otimes_{k(g)} L$  is identified with  $B^{r(N-r)}$ .

*Proof.* Let  $\{q_1, \dots, q_N\}$  be the standard basis for the space of global sections of  $\mathcal{Q}$ , and consider the canonical map  $k^N \otimes_k \mathcal{O}_{Gr} \rightarrow \mathcal{Q}$ . Without loss of generality, we may assume that the images of  $q_1, \dots, q_r$  generate  $\mathcal{Q} \otimes k(g)$ . Let  $U \subset Gr$  be the open set where the  $k$ -vectors  $q_1, \dots, q_r$  give a basis for  $\mathcal{Q}$ . Then  $U \cong A_k^{r(N-r)}$ . If  $\{X_{ij} | 1 \leq i \leq r, 1 \leq j \leq N-r\}$  are the coordinates on this affine space, where the basis of  $\mathcal{O}_U^N$  is  $e_1, \dots, e_r, f_1, \dots, f_{N-r}$ , and  $\psi(f_j) = \sum_i X_{ij} \psi(e_i)$  for the map  $\psi: k^N \otimes \mathcal{O}_U \rightarrow \mathcal{Q}|_U$ , then  $dX_{ij}$  form a basis for the free  $\mathcal{O}_U$ -module  $\Omega_{U/k}^1$ .

Let the variables on  $\mathbf{A}^{rN}$  be  $Y_{11}, \dots, Y_{rr}, Z_{11}, \dots, Z_{r(N-r)}$ ; then by equation (1),

$$v_i = \sum_{j=1}^r Y_{ij} \psi(e_i) + \sum_{l=1}^{N-r} Z_{il} \sum_{m=1}^r X_{ml} \psi(e_m)$$

and therefore (see equation (2))

$$C_0(Gr)|_U = \{\mathbf{q} \in \mathbf{A}^{rN} \times U \mid \det(t_{ij}) \otimes k(\mathbf{q}) = 0\}$$

where  $t_{ij} = (Y_{ij} + \sum_{l=1}^{N-r} Z_{il} X_{jl})$ ; this implies that for any  $g \in Gr$ , we have

$$C_0(Gr)_g = \{\mathbf{p} \in \mathbf{A}^{rN} \mid (\det(t_{ij}) \otimes k(g)) \otimes k(\mathbf{p}) = 0\} \subseteq \mathbf{A}_{k(g)}^{rN}.$$

From the formula for  $t_{ij}$ , it is clear that  $t_{ij}$ ,  $1 \leq i, j \leq r$  and  $Z_{kl}$ ,  $1 \leq k \leq r$ ,  $1 \leq l \leq N-r$  together a system of coordinates on  $\mathbf{A}^{rN}$ .

Let

$$T_{ij} = (i, j) \text{th cofactor of } [t_{ij}].$$

Then by a simple computation one can see that

$$d(\det(t_{ij})) = \sum_{i,j} T_{ij} d(t_{ij}).$$

But

$$d(t_{ij}) = \sum_{l=1}^{N-r} Z_{il} dX_{jl}.$$

Therefore

$$\begin{aligned} d(\det(t_{ij})) &= \sum_{i,j} \sum_l T_{ij} Z_{il} dX_{jl} \\ &= \sum_{j=1}^r \sum_{l=1}^{N-r} \left( \sum_{i=1}^r T_{ij} Z_{il} \right) (dX_{jl}). \end{aligned}$$

Hence Image  $(\bar{d} \otimes 1)$  is spanned by the element

$$\xi := \sum_{j=1}^r \sum_{l=1}^{N-r} \left( \sum_{i=1}^r T_{ij} Z_{il} \right) (dX_{jl} \otimes 1_{k(g)}).$$

For the sake of brevity, from now onwards we denote

1. the basis  $\{d(X_{jl}) \otimes 1_L\}$  by  $\{\tilde{e}_i \mid 1 \leq i \leq r(N-r)\}$
2. the coordinates of  $\xi$ , namely  $\{\sum_{k=1}^r T_{kj} Z_{kl}\}$ , by  $\{\tilde{Y}_i \mid 1 \leq i \leq r(N-r)\}$ .

Let  $\gamma: B^{r(N-r)} \rightarrow B^m$  be a nonzero  $B$ -linear map and let  $[b_{ij}]_{r(N-r) \times m}$  be the matrix of  $\gamma$  corresponding to the standard basis  $\{\tilde{e}_i\}_i$  of  $B^{r(N-r)}$  and some basis say  $\{f_j\}_j$  of  $B^m$ . The map  $\gamma: B^{r(N-r)} \rightarrow B^m$  is given by  $\sum a_i e_i \rightarrow \sum_{i,j} a_i b_{ij} f_j$ . Therefore

$$\begin{aligned} \gamma\left(\sum \tilde{Y}_i e_i\right) &= 0 \Leftrightarrow \left(\sum_i \tilde{Y}_i b_{ij}\right) f_j = 0 \quad \text{for each } j \\ &\Leftrightarrow \sum_i \tilde{Y}_i b_{ij} = 0 \quad \text{in } B \text{ for each } j. \end{aligned}$$

Since the map  $\gamma$  is not zero, at least one of the  $b_{ij}$  is not zero. Since  $B = L[\{t_{ij}\}_{r \times r}, \{Z_{il}\}_{r \times (N-r)}] / (\det(t_{ij}))$  to prove the lemma, it is enough to prove that for any  $i_0 \in \{1, 2, \dots, r(N-r)\}$ , there exists  $(\alpha, \gamma) = ((\alpha_{ij})_{r \times r}, (\gamma_{il})_{r \times (N-r)})$  in  $L^N$  such that

1.  $\det(\alpha_{ij}) = 0$  (i.e.,  $(\alpha, \gamma) \in \text{Spec}(B)$ )
2.  $(\alpha, \gamma)$  satisfies the equations  $\tilde{Y}_i = 0$  for  $i \neq i_0$
3.  $(\alpha, \gamma)$  does not satisfy the equation  $\tilde{Y}_{i_0} = 0$

If  $\tilde{Y}_{i_0} = T_{1k}Z_{1l} + T_{2k}Z_{2l} + \dots + T_{rk}Z_{rl}$  then to prove this it is enough to find  $\alpha = (\alpha_{ij}) \in L^{r \times N}$  such that i)  $\det(\alpha_{ij}) = 0$ , ii)  $(\alpha_{ij})$  satisfies the equation  $T_{ij} = 0$  for all  $(i, j) \neq (1, k)$  but iii) does not satisfy the equation  $T_{1k} = 0$ ; as we can then choose  $(\gamma) = E_{1l}$  where  $E_{1l}$  denotes the elementary matrix with  $(1, l)$ th entry 1 and other entries equal to zero. Let

$$(\alpha_{ij}) = \begin{bmatrix} 0 & 0 & 0 \\ I_{k-1} & 0 & 0 \\ 0 & 0 & I_{r-k} \end{bmatrix}$$

where  $I_{k-1}$  and  $I_{r-k}$  denote the identity matrices of size  $(k-1)$  and  $(r-k)$  respectively. If  $A_{ij}$  denotes the  $(i, j)$ th cofactor of  $(\alpha_{ij})$  then it is clear that  $A_{1k} = \pm 1$  and  $A_{ij} = 0$  for  $(i, j) \neq (1, k)$ . Hence the lemma.  $\square$

### 3. Height

#### PROPOSITION 1

Let  $(R, \mathfrak{m})$  be a local ring containing an infinite perfect field  $k$  such that  $\hat{R}$  is an equidimensional ring. Let  $I \subset R$  be an ideal of  $R$  of height  $r$  such that  $I = \langle g_1, \dots, g_N \rangle$  and let  $\mathfrak{F}_\alpha = \langle \sum \alpha_{1j} g_j, \dots, \sum \alpha_{rj} g_j \rangle$ , where  $\alpha_{ij} \in k$ . Then for general  $\alpha = (\alpha_{ij}) \in k^{rN}$ , the ideal  $\mathfrak{F}_\alpha$  is of height  $r$ .

*Proof.* Since  $R \rightarrow \hat{R}$  is a faithfully flat extension, we have  $\text{ht}(I) = \text{ht}(I\hat{R})$  and  $\text{ht}(\mathfrak{F}_\alpha) = \text{ht}(\mathfrak{F}_\alpha \hat{R})$ . Therefore we can assume without loss of generality that  $R$  is a complete local ring. Moreover if the result is true when  $k \cong R/\mathfrak{m}$  is a coefficient field, then it is true for arbitrary infinite perfect field  $k \subseteq R$ , since such a field is contained in a coefficient field of  $R$  (see [M]). Therefore we can also assume that  $k = R/\mathfrak{m}$ .

If  $\dim R = m$  then there exist  $h_1, \dots, h_{m-r}$  in  $\mathfrak{m}$  such that the ideal  $(g_1, \dots, g_N, h_1, \dots, h_{m-r})$  is  $\mathfrak{m}$ -primary. Let  $R' = k[[g_1, \dots, g_N, h_1, \dots, h_{m-r}]]$  be the complete  $k$ -subalgebra of  $R$  generated by  $g_1, \dots, g_N, h_1, \dots, h_{m-r}$ . Then  $R$  is finite over  $R'$ , so that the ring  $R/(h_1, \dots, h_{m-r})$  is a finite  $R'/(h_1, \dots, h_{m-r})$ -module.

Let  $S = R'/(h_1, \dots, h_{m-r}) = k[[g_1, \dots, g_N]]$  and let  $S_\alpha = k[[f_1, \dots, f_r]]$  be the complete subring of  $S$ , where  $f_i = \sum \alpha_{ij} g_j$  and  $(\alpha_{ij}) \in k^{rN}$ . If the canonical map of associated graded rings  $i_\alpha: \text{gr}(S_\alpha) \rightarrow \text{gr}(S)$  makes  $\text{gr}(S)$  a finite (graded)  $\text{gr}(S_\alpha)$ -module for some  $\alpha$  in  $k^{rN}$ , then

$$\dim \left( \frac{\text{gr}(S)}{i_\alpha(\bar{f}_1, \dots, \bar{f}_r) \text{gr}(S)} \right) = \dim \left( \frac{\bigoplus_{t=0}^{\infty} I^t}{I^{t+1} + (\bar{f}_1, \dots, \bar{f}_r) I^{t-1}} \right) = 0,$$



where  $I = (g_1, \dots, g_N)$  is the maximal ideal of  $S$ . But

$$\text{gr}\left(\frac{k[[g_1, \dots, g_N]]}{(f_1, \dots, f_r)}\right) = \bigoplus_{i=0}^{\infty} \frac{I^i + (f_1, \dots, f_r)}{I^{i+1} + (f_1, \dots, f_r)} = \bigoplus_{i=0}^{\infty} \frac{I^i}{(I^i \cap (f_1, \dots, f_r)) + I^{i+1}}$$

which is a quotient ring of  $\text{gr}(S)/i_{\alpha}(\bar{f}_1, \dots, \bar{f}_r)$ . Therefore

$$\dim(k[[g_1, \dots, g_N]]/(f_1, \dots, f_r)) = 0.$$

Hence  $(f_1, \dots, f_r)S$  is  $(g_1, \dots, g_N)$ -primary, which implies

$$(f_1, \dots, f_r, h_1, \dots, h_{m-r})R' \text{ is } (g_1, \dots, g_N, h_1, \dots, h_{m-r})\text{-primary,}$$

which in turn implies  $(f_1, \dots, f_r, h_1, \dots, h_{m-r})R$  is  $\mathfrak{m}$ -primary. Therefore  $\text{ht}(f_1, \dots, f_r) = r$ . Hence to prove that  $\text{ht}(\mathfrak{F}_{\alpha}) = r$  for general  $(\alpha) \in k^{r \cdot N}$ , it is enough to prove that  $\text{gr}(S)$  is finite over  $\text{gr}(S_{\alpha})$  for general  $(\alpha) \in k^{r \cdot N}$ . This is essentially the Noether normalisation theorem; we sketch the proof below.

Let  $\text{Proj } k[X_1, \dots, X_N, Y] = \mathbf{P}^N = \mathbf{P}^{N-1} \amalg \mathbf{A}^N$  where the map  $\mathbf{A}^N \hookrightarrow \mathbf{P}^N$  is given by  $(x_1, \dots, x_N) \mapsto (x_1 : \dots : x_N : 1)$  and  $\mathbf{P}^{N-1} \hookrightarrow \mathbf{P}^N$  is given by  $(x_1 : \dots : x_N) \mapsto (x_1 : \dots : x_N : 0)$ .

Let  $F_1 = \sum \alpha_{1i} X_i, \dots, F_r = \sum \alpha_{ri} X_i$ , where  $1 \leq i \leq N$  and let  $\mathbf{L}_{\alpha} = V(F_1, \dots, F_r) \subseteq \mathbf{P}^{N-1}$  be the corresponding linear subspace. Let  $k^{r \cdot N} = \text{Spec } k[\{Y_{ij}\}_{r \times N}]$ . Let  $h(Y_{ij})$  be an  $r \times r$  minor of the matrix of indeterminates  $[Y_{ij}]_{r \times N}$  and let  $U = k^{r \cdot N} \setminus V(h)$  which is a nonempty open set of  $k^{r \cdot N}$ . Then for  $(\alpha_{ij}) \in U$ , we have  $\dim \mathbf{L}_{\alpha} = (N-1) - r$ .

Consider the following diagram of linear projections.

$$\begin{array}{ccccc} \mathbf{A}^N & \hookrightarrow & \mathbf{P}^N \setminus \mathbf{L}_{\alpha} & \hookrightarrow & \mathbf{P}^{N-1} \setminus \mathbf{L}_{\alpha} \\ \downarrow & & \downarrow \pi_{\alpha} & & \downarrow \\ \mathbf{A}^r & \hookrightarrow & \mathbf{P}^r & \hookrightarrow & \mathbf{P}^{r-1} \end{array}$$

The map  $\pi_{\alpha}: \mathbf{P}^N \setminus \mathbf{L}_{\alpha} \rightarrow \mathbf{P}^r$  is given by  $(x_1 : \dots : x_N : y) \mapsto (F_1(x) : \dots : F_r(x) : y)$ . Then  $\pi_{\alpha}^{-1}(\mathbf{A}^r) = \mathbf{A}^N$ . Let  $k[X_1, \dots, X_N] \twoheadrightarrow \text{gr}(S)$  be the surjective graded homomorphism of rings sending  $X_i \mapsto \tilde{g}_i$  (where  $\tilde{g}_i$  is the image in  $I/I^2$  of  $g_i$ ), then the scheme  $X := \text{Spec } \text{gr}(S)$  is a closed subset of  $\mathbf{A}^N$  and is of dimension  $r$ . Let  $\bar{X} \subset \mathbf{P}^N$  be the closure of  $X$ .

If  $\bar{X} \cap \mathbf{L}_{\alpha} = \emptyset$  then  $\pi_{\alpha}: \bar{X} \rightarrow \mathbf{P}^r$  is proper, which implies  $\pi_{\alpha}: \pi_{\alpha}^{-1}(\mathbf{A}^r) \cap \bar{X} \rightarrow \mathbf{A}^r$  is proper, hence finite. Since  $\pi_{\alpha}^{-1}(\mathbf{A}^r) \cap X = \mathbf{A}^N \cap \bar{X} = X$ , this implies that the ring  $\text{gr}(S)$  is finite over  $\text{gr}(S_{\alpha})$  if  $\bar{X} \cap \mathbf{L}_{\alpha} = \emptyset$ . But  $\bar{X} \cap \mathbf{L}_{\alpha} = \emptyset \Leftrightarrow (\bar{X} \setminus X) \cap \mathbf{L}_{\alpha} = \emptyset$ .

Since  $\dim(\bar{X} \setminus X) = r-1$  and  $\dim \mathbf{L}_{\alpha} = N - (r+1)$ , we claim that for general  $(\alpha)$ , the set  $\mathbf{L}_{\alpha} \cap \bar{X} = \emptyset$ . To see this, consider the incidence variety  $\Gamma \subset (\bar{X} \setminus X) \times \mathbf{A}^{r \cdot N}$ ,

$$\Gamma = \{(x, \alpha) | x \in \mathbf{L}_{\alpha}\}.$$

For each  $x \in \bar{X} \setminus X$ , the fibre  $\Gamma_x$  of  $\Gamma$  over  $x$  is a linear subspace  $\mathbf{A}^{r(N-1)} \subset \mathbf{A}^{r \cdot N}$ , so that  $\dim \Gamma = \dim \bar{X} \setminus X + r(N-1) = rN - 1 < \dim \mathbf{A}^{r \cdot N}$ . Hence  $\Gamma \rightarrow \mathbf{A}^{r \cdot N}$  is not dominant.  $\square$

#### 4. Smoothness on $D(I)$

For a local ring  $R$ , let  $\text{reg}(R)$  be the set of regular points, and for an ideal  $I$  in  $R$ , let  $D(I) = \text{Spec}(R) \setminus \text{Spec}(R/I)$ .

**Theorem 2.** Let  $(R, \mathbf{m})$  be a local Nagata ring containing an infinite perfect field  $k$  and let  $I \subset R$  be an ideal. Then there exist generators  $g_1, \dots, g_N$  of  $I$  such that if  $\mathfrak{F}_\alpha$  denotes the ideal  $\langle \sum_j g_j \alpha_{1j}, \dots, \sum_j g_j \alpha_{rj} \rangle$ , where  $r$  is some integer, then for general  $(\alpha_{ij}) \in k^{rN}$ , the ring  $R/\mathfrak{F}_\alpha$  is regular at all primes of  $D(I) \cap \text{reg}(R)$ .

*Proof.* Let  $\{a_1, \dots, a_k\}$  be a set of generators of  $I$  and let  $\{g_1, \dots, g_N\}$  be the set  $\{a_1, \dots, a_k, x_1 a_1, x_1 a_2, \dots, x_l a_k\}$  where  $\{x_1, \dots, x_l\}$  is a set of generators of  $\mathbf{m}$ , and  $N = k + kl$ .

*Reduction 1.* One can assume that  $R$  is a complete domain and  $k = R/\mathbf{m}$ .

Let  $S_\alpha = R/\mathfrak{F}_\alpha$  and let  $K = R/\mathbf{m}$ . Since  $R$  is Nagata, and the maps  $R \rightarrow \hat{R}$  and  $S_\alpha \rightarrow \hat{S}_\alpha$  are faithfully flat, for the induced maps  $\eta: \text{Spec } \hat{R} \rightarrow \text{Spec } R$  and  $\eta_\alpha: \text{Spec } \hat{S}_\alpha \rightarrow \text{Spec } S_\alpha$  we have

$$\eta_\alpha(D(I\hat{S}_\alpha)) = D(IS_\alpha) \text{ and } \eta(\text{reg}(\hat{R})) = \text{reg}(R).$$

Further, if  $\hat{S}_\alpha$  is regular on  $D(I\hat{S}_\alpha)$  then  $S_\alpha$  is regular on  $D(IS_\alpha)$  (see [M]). Moreover, if on  $D(I\hat{S}_\alpha)$ ,  $\hat{S}_\alpha$  is regular for general  $(\alpha) \in K^{rN}$  then it is regular for general  $(\alpha) \in k^{rN}$ .

Since  $\text{reg}(R) \subset \cup_i \text{reg}(R_i)$ , where  $\text{Spec } R_i$  are the irreducible components of  $\text{Spec } R$ , it is enough to prove the result for each  $R_i$ . Hence we may also assume that  $R$  is a domain. This justifies the first reduction.

Let  $M = R^{\oplus r} \oplus D_k(R)$ , where  $D_k(R)$  is the universal finite differential module of  $k$ -derivations (see [K]). Let

$$\text{rank } D_k(R) = d.$$

Let

$$\varphi: R^{\oplus r} \rightarrow R^{\oplus r} \oplus D_k(R)$$

be defined by  $(t_1, \dots, t_r) \mapsto (t_1, \dots, t_r, \sum dt_i)$ . Let  $e_1, \dots, e_r$  be the standard basis for  $R^{\oplus r}$ , and let  $h_{ij} \in M$  be defined by

$$h_{ij} = \varphi(g_j e_i) = \varphi(0, \dots, 0, g_j, 0, \dots, 0).$$

Let  $N_i$  be the submodule generated by  $\langle h_{i1}, \dots, h_{iN} \rangle$  and let  $N = \sum_1^r N_i$ . Then we have

*Lemma.* With the notations of the above paragraph we have  $N_{\mathbf{p}} = M_{\mathbf{p}}$ , for all  $\mathbf{p} \in D(I)$ .

*Proof of the lemma.* If  $\mathbf{p} \in D(I)$ , then there exists some  $a_i$  such that  $a_i \notin \mathbf{p}$ . Now for given  $x_j \in \mathbf{m}$ , choose  $g_{j1} = x_j a_i$  and  $g_{j2} = a_i$ . Then

$$\begin{aligned} h_{1j1} - x_j h_{1j2} &= (x_j a_i, 0, \dots, 0, a_i dx_j + x_j da_i) - x_j(a_i, 0, \dots, 0, da_i) \\ &= a_i(0, \dots, 0, dx_j) \end{aligned}$$

Therefore  $\langle \{h_{ij}\}_{i,j} \rangle_{\mathbf{p}} \supseteq \langle dx_1, \dots, dx_l \rangle$ . But  $\{dx_j\}_j$  generates  $D_k(R)$ , hence  $\langle \{h_{ij}\}_{i,j} \rangle_{\mathbf{p}} \supseteq D_k(R)_{\mathbf{p}}$ . Therefore for  $a_i$ , the element  $da_i \in \langle \{h_{ij}\}_{i,j} \rangle_{\mathbf{p}}$ . But  $h_{ij2} - da_i = (0, \dots, a_i, \dots, 0)$ , where  $a_i$  is at the  $i$ th place; which implies  $\langle \{h_{ij}\}_{i,j} \rangle_{\mathbf{p}} \supseteq R_{\mathbf{p}}^{\oplus r}$ . Hence the lemma.  $\square$

Consider the following sequence of  $R$ -modules

$$0 \rightarrow N_i \xrightarrow{f_i} N \xrightarrow{\eta} I^{\oplus r-1} \rightarrow 0 \quad (5)$$

where  $f_i$  is the natural inclusion as a submodule and the surjective map

$\eta_i: N \rightarrow I^{\oplus r-1}$  is given by

$$(a_1, \dots, a_r, d) = (a_1, \dots, \hat{a}_i, \dots, a_r),$$

where  $d \in D_k(R)$  and  $\hat{a}_i$  means that  $a_i$  is deleted. We note that  $f_i(N_i) \subseteq \ker \eta_i$ .

*Claim.* The above sequence is a split short exact sequence.

*Proof of the claim.* Define  $h_i: I^{\oplus r-1} \rightarrow N$  as

$$(a_1, \dots, a_{r-1}) \mapsto \left( a_1, \dots, -\sum_{j=1}^{r-1} a_j, a_i, \dots, a_{r-1}, 0 \right),$$

It is easy to see that  $\eta_i \circ h_i = \text{Id}$ . Therefore to prove the claim it is enough to prove that  $h_i(I^{\oplus r-1}) + N_i = N$ ; which follows easily as the set  $\{\varphi(g_j e_k) = (g_j e_k, dg_j) \mid 1 \leq k \leq r, 1 \leq j \leq N\}$  generates  $N$  and  $(g_j e_k, dg_j) = (g_j e_k - g_j e_i, 0) + (g_j e_i, dg_j) \in h_i(I^{\oplus r-1}) + N_i$ .

In particular  $N$  is of rank  $r + d$  and  $N_i$  is free of rank  $d + 1$ .

Consider the ring  $S = \text{Sym}_R(\wedge^{r+d} N)/(\text{torsion})$ . Let  $X = \text{Proj } S$ , and let  $\pi: \text{Proj } S \rightarrow \text{Spec } R$  be the canonical map (if  $\tilde{N}$  is the coherent sheaf corresponding to  $N$  on  $\text{Spec}(R)$  then  $\pi$  is the Nash blow up for  $\tilde{N}$  on  $\text{Spec}(R)$ ).

By lemma 2,  $S$  is a domain and  $N' := (N \otimes S)/(\text{torsion})$  yields a locally free sheaf  $\tilde{N}'$ , of rank  $r + d$ , on  $X$ . Consider the closed subscheme  $Y = \pi^{-1}(m)$  of  $X$  with the reduced scheme structure, so that  $Y$  is a projective variety over  $k$ . Then  $\mathcal{E} := \tilde{N}' \otimes_{\mathcal{O}_X} \mathcal{O}_Y$  is a locally free  $\mathcal{O}_Y$ -module of rank  $(r + d)$  generated by  $\{\tilde{h}_{ij} \mid 1 \leq i \leq r, 1 \leq j \leq N\}$ , where  $\tilde{h}_{ij} := h_{ij} \otimes_{\mathcal{O}_X} 1$ . Similarly if  $N'_i$  denotes the module  $(N_i \otimes S)/(\text{torsion})$  then the sheaf  $\mathcal{E}_i := \tilde{N}'_i \otimes_{\mathcal{O}_X} \mathcal{O}_Y$  is a locally free  $\mathcal{O}_Y$ -module (as it is a direct summand of  $\mathcal{E}$ ), which is of rank  $d + 1$ , generated by  $\{\tilde{h}_{ij} \mid 1 \leq j \leq N\}$ . Moreover, the subsheaf  $\cap_i \mathcal{E}_i$  is locally free of rank  $d$ .

*Reduction 2.* To prove the theorem it is sufficient to prove that the set  $\{\tilde{h}_i = \sum_{j=1}^N h_{ij} \alpha_{ij} \mid 1 \leq i \leq r\}$  generates an  $r$ -fold basic  $\mathcal{O}_Y$ -submodule of  $\mathcal{E}$ , for general  $(\alpha_{ij}) \in k^{rN}$ .

For, by lemmas 3 and 4, this will imply that  $\langle \{\sum h_{ij} \alpha_{ij} \mid 1 \leq i \leq r\} \rangle$  is an  $r$ -fold basic submodule of  $N$ . Therefore, by the lemma above, it is an  $r$ -fold basic submodule of  $M$ , at all  $\mathfrak{p} \in D(I)$ . Therefore if  $\{\sum_j \alpha_{ij} g_j \mid 1 \leq i \leq r\} \subseteq \mathfrak{p}$  then  $\langle \sum \alpha_{1j} d(g_j), \dots, \sum \alpha_{rj} d(g_j) \rangle$  is an  $r$ -fold basic submodule of  $D_k(R)$  at  $\mathfrak{p}$ . Hence by lemma 7, the ring  $R/\mathfrak{F}_x$  is regular at  $\mathfrak{p}$ .

Moreover, it is enough to prove the results for each irreducible component of  $Y$ . Hence we can assume  $Y$  is an irreducible projective  $k$ -variety. Let  $V_i$  denote an  $n$  dimensional  $k$ -vector space for  $1 \leq i \leq r$  and let  $V_i^*$  denote the dual of  $V_i$ . Then  $Z := V_1^* \times \dots \times V_r^* \times Y = k^{rN} \times Y$  is a  $k$ -variety with natural projections  $q: Z \rightarrow Y$  and  $p_i: Z \rightarrow V_i^*$ . Let  $\psi_i: V_i \otimes_k \mathcal{O}_Z \rightarrow q^* \mathcal{E}_i$  be the map taking the basis  $\{\tilde{e}_{i1} \otimes 1, \dots, \tilde{e}_{iN} \otimes 1\}$  of  $V_i \otimes_k \mathcal{O}_Z$  onto the set of generators  $\{\tilde{h}_{ij} \mid 1 \leq j \leq N\}$  of  $q^* \mathcal{E}_i$ . Extend this map to  $\psi: \oplus_i (V_i \otimes_k \mathcal{O}_Z) \rightarrow q^* \mathcal{E}$  such that  $\psi|_{V_i \otimes_k \mathcal{O}_Z} \equiv \psi_i$ , where we regard  $q^* \mathcal{E}_i$  as a subsheaf of  $q^* \mathcal{E}$ . We define

$$\Gamma = \{(\alpha_{ij}, \mathfrak{p}) \in Z \mid (\oplus_i V_i \otimes_k \mathcal{O}_Z) \otimes_{\mathcal{O}_Z} k(\alpha_{ij}, \mathfrak{p}) \xrightarrow{\psi \otimes \text{Id}} q^* \mathcal{E} \otimes_{\mathcal{O}_Z} k(\alpha_{ij}, \mathfrak{p}).$$

is of rank  $< r \}$ .

Let  $\mathcal{O}_{V_r^*} = k[\tilde{x}_{i1}, \dots, \tilde{x}_{iN}]$  where  $\tilde{x}_{ij}$  are dual to the basis  $\{\tilde{e}_{ij}\}_{j=1}^N$  of  $V_i$ . Take  $s_i = \sum_j \tilde{e}_{ij} \otimes \tilde{x}_{ij} \in V_i \otimes \mathcal{O}_{V_r^*}$ ; then  $\psi_i(p_i^*(s_i)) \in q^* \mathcal{E}_i$ , where  $p_i^*: V_i \otimes \mathcal{O}_{V_r^*} \rightarrow V_i \otimes \mathcal{O}_Z$  is the canonical map. Hence

$$\psi_1(p_1^*(s_1)) \wedge \dots \wedge \psi_r(p_r^*(s_r)) \in q^*(\wedge^r \mathcal{E})$$

and the zero locus of this section is precisely  $\Gamma$ . Note that  $s_i$  is well defined as it is independent of the choice the basis for  $V_i$  and corresponding dual basis for  $V_i^*$ . To prove the result it is enough to prove that  $p(\Gamma)$  is a proper subset of  $V_1^* \times \dots \times V_r^* = k^{rN}$ , where  $p: Z \rightarrow V_1^* \times \dots \times V_r^*$  is the natural projection.

For this we compute  $\dim(\Gamma_y)$  for  $y \in Y$ , where  $\Gamma_y = q^{-1}(y)$  denotes the fibre  $\Gamma$  over  $y \in Y$ . Consider the map  $\tilde{\psi}: \oplus_i (V_i \otimes_k \mathcal{O}_Z) \otimes_{\mathcal{O}_Y} k(y) \rightarrow q^* \mathcal{E} \otimes_{\mathcal{O}_Y} k(y)$ , which is

$$\tilde{\psi}: (\oplus_i V_i) \otimes_k S(V_1^* \oplus \dots \oplus V_r^*) \otimes k(y) \rightarrow W \otimes_k S(V_1^* \oplus \dots \oplus V_r^*),$$

where  $W = \mathcal{E} \otimes_{\mathcal{O}_Y} k(y)$  is a  $k(y)$ -vector space of dimension  $(r+d)$ . Now the section

$$\tilde{\psi}(s_1) \wedge \dots \wedge \tilde{\psi}(s_r) \in (\wedge^r W) \otimes S(V_1^* \oplus \dots \oplus V_r^*).$$

For a given basis of  $W$ , we get a basis of  $(\wedge^r W)$ , so we get components of  $\tilde{\psi}(s_1) \wedge \dots \wedge \tilde{\psi}(s_r)$  which are elements of  $S(V_1^* \oplus \dots \oplus V_r^*) \otimes k(y)$ . The ideal generated by these components in  $S(V_1^* \oplus \dots \oplus V_r^*) \otimes k(y)$  is the ideal of the fibre  $\Gamma_y$  regarded as a subvariety of  $(V_1^* \times \dots \times V_r^*)_{k(y)}$ . Let  $W_i = \mathcal{E}_i \otimes_Y k(y)$  and let  $W' = \cap W_i$ . It is clear that  $\dim W' = d$  and  $\psi(V_i \otimes k(y)) = W_i$ . We choose a basis  $f_1, \dots, f_d$  of  $W'$  and choose  $g_i \in W_i \setminus W'$  such that  $g_1, \dots, g_r, f_1, \dots, f_d$  is a basis for  $W$ . For each  $V_i \otimes k(y) = W'_i \oplus W''_i$  (where  $W'_i \cong W_i$  via a splitting of the surjective map  $V_i \otimes k(y) \rightarrow W_i$ ), choose a basis  $e_i, e_{i1}, \dots, e_{id}$  of  $W_i$  and extend it to a basis  $e_i, e_{i1}, \dots, e_{id}, e_{id+1}, \dots, e_{i(N-1)}$  of  $V_i \otimes k(y)$ . Let  $x_i, y_{i1}, \dots, y_{id}, y_{id+1}, \dots, y_{i(N-1)}$  be the corresponding dual basis of  $V_i^*$ . Then

$$s_i = e_i \otimes x_i + \sum_j^{N-1} e_{ij} \otimes y_{ij}$$

which implies that

$$\tilde{\psi}(s_i) = g_i \otimes x_i + \sum_{j=1}^d f_j \otimes y_{ij} \in W \otimes S(V_1^* \oplus \dots \oplus V_r^*).$$

Therefore the ideal of  $\Gamma_y$  in  $S(V_1^* \oplus \dots \oplus V_r^*) \otimes k(y)$  is generated by  $r \times r$  minors of matrix

$$\begin{pmatrix} x_1 & 0 & y_{11} & \dots & y_{1d} \\ & \ddots & \vdots & & \vdots \\ 0 & x_r & y_{r1} & \dots & y_{rd} \end{pmatrix}$$

Hence, by lemma 5, we have  $\dim \Gamma_y = (r-1)(d+1) + Nr - (r+rd)$ .

Let  $\Gamma = \cup \Gamma_i$ , where  $\Gamma_i$  denotes an irreducible component of  $\Gamma$ . Then

$$\begin{aligned} \dim \Gamma_i &\leq (r-1)(d+1) + Nr - (r+rd) + \dim Y \\ &= Nr + \dim Y - (d+1) \quad \text{for all } i. \end{aligned}$$

*Example 3.1.* Let  $k$  be some infinite field of characteristic  $p > 0$ . Let  $I = (X^p Z^p, X^p W^p, Y^p Z^p, Y^p W^p) \subset k[[X, Y, Z, W]]$ , with chosen generators  $g_1 = X^p Z^p$ ,  $g_2 = X^p W^p$ ,  $g_3 = Y^p Z^p$ ,  $g_4 = Y^p W^p$ . Then for any  $\alpha = (\alpha_{ij})$ , where  $\alpha_{ij} \in k$ ,  $1 \leq i \leq 2$ ,  $1 \leq j \leq 4$ , let  $\beta_{ij}^p = \alpha_{ij}$ . Then

$$\mathfrak{F}_\alpha = ((\beta_{11} g_1 + \beta_{12} g_2 + \beta_{13} g_3 + \beta_{14} g_4)^p, (\beta_{21} g_1 + \beta_{22} g_2 + \beta_{23} g_3 + \beta_{24} g_4)^p).$$

Clearly  $R/\mathfrak{F}_\alpha$  is not regular at any (non-zero) localisation. Now  $I$  is not a set-theoretic complete intersection, since the punctured spectrum of  $R/I$  is not connected (see [H2], Theorem 2.2). Hence  $\text{Spec } R/\mathfrak{F}_\alpha$  has non-empty intersection with  $D(I)$ .

*Remark.* By giving a few more technical arguments in the above theorem, we can replace the assumption ‘ $R$  is a local Nagata ring’ by ‘ $R$  is a Noetherian local ring’.

## 5. A Bertini theorem in codimension 1 for modules

We consider the following situation. Let  $(R, m)$  be a local Nagata ring containing an infinite perfect field  $k$  of arbitrary characteristic. Let  $M$  be a finitely generated  $R$ -module of rank  $r$  (i.e.,  $M$  is free of rank  $r$ , at each minimal prime  $\mathfrak{p}$  of  $R$ ). Let  $U_R$  be the open subset consisting of points of  $\text{Spec } (R)$  on which  $M$  is free. Let

$$V_R = \{\mathfrak{p} \in U_R \mid R_{\mathfrak{p}} \text{ is regular and of } \dim \leq 1\}.$$

Let  $\{g_1, \dots, g_N\}$  denote a set of generators of  $M$ . Let  $\psi_i: R \rightarrow M$  be the  $R$ -linear maps given by

$$1 \mapsto \alpha_{i1} g_1 + \dots + \alpha_{iN} g_N \quad \text{for } 1 \leq i \leq r,$$

where  $\alpha = (\alpha_{ij}) \in k^{rN}$ . Let  $\psi$  be the map  $\psi_1 \wedge \dots \wedge \psi_r: R \rightarrow \wedge^r M$ . Consider the ideal

$$T_\alpha = \psi^*((\wedge^r M)^*) \subseteq R, \quad (6)$$

where  $\psi^*: (\wedge^r M)^* \rightarrow R$  is the dual of  $\psi$ . If for some  $\mathfrak{p} \in \text{Spec}(R)$  the module  $M_{\mathfrak{p}}$  is generated by  $r$  elements, say  $g_1, \dots, g_r$ , and if  $g_k = a_{1k} g_1 + \dots + a_{rk} g_r$ , where  $k \geq r+1$  and  $a_{ik} \in R_{\mathfrak{p}}$ , then  $T_{\alpha\mathfrak{p}} = T_\alpha R_{\mathfrak{p}}$  is generated by the element  $\det([\alpha_{ij} + \sum_{k \geq r+1} \alpha_{ik} a_{jk}]_{r \times r})$ .

We first state a result in characteristic 0.

**Theorem 3.** *Suppose  $k$  is a field of characteristic 0. In the above situation, for an arbitrary choice of the generators  $g_1, \dots, g_N$  of  $M$ , the ring  $R/T_\alpha$  is regular on  $V_R$  for general  $\alpha \in k^{rN}$ .*

In arbitrary characteristic, we have a slightly weaker result.

**Theorem 4.** *Let  $k$  be a perfect field of arbitrary characteristic. In the above situation, for a suitable choice of generators  $\{g_1, \dots, g_N\}$  of  $M$  and for general  $\alpha = (\alpha_{ij}) \in k^{rN}$ , the ring  $R/T_\alpha$  is regular on  $V_R$ .*

*The generators  $\{g_1, \dots, g_N\}$  may be chosen as follows: if  $\{m_1, \dots, m_k\}$  is a set of*

generators of  $M$ , then it suffices that  $\{g_1, \dots, g_N\} \supset \{m_1, \dots, m_k, m\}$  for a suitable  $m \in M$ ; if  $\{x_1, \dots, x_n\}$  is a set of generators for the maximal ideal  $\mathfrak{m}$  of  $R$ , we may take  $m = x_i m_j$  for suitable  $i, j$ .

*Remark.* By lemma 6, to prove the assertion that the ring  $R/T_\alpha$  is regular on  $V_R$ , one can assume that  $R$  is a complete local domain. Moreover, if the statement is true for the field  $R/\mathfrak{m}$  then it is true for the field  $k(\subset R/\mathfrak{m})$ . Hence we can also assume that  $k = R/\mathfrak{m}$ . Finally, if  $\dim R \leq 1$ , the result is trivial, so we may assume  $\dim R > 1$ .

*Proof of theorem 3.* Let  $\tilde{M}$  be the coherent sheaf on  $\text{Spec}(R)$  corresponding to  $M$ . Consider the integral scheme  $X_1$  which is the Nash blow up of  $\text{Spec} R$  for  $\tilde{M}$ , and the canonical map  $\pi_1: X_1 \rightarrow \text{Spec}(R)$ . Then  $\mathcal{M} := \pi_1^* \tilde{M} / (\text{torsion})$  is a locally free  $\mathcal{O}_{X_1}$ -module of rank  $r$ , by lemma 2. Let  $Y_1 = \pi_1^{-1}(\mathfrak{m})_{\text{red}}$  be the reduced closed subscheme of  $X_1$ .

Let  $\pi_2: X_2 \rightarrow X_1$  be the blow-up map with respect to the coherent ideal sheaf  $\mathcal{I}_{Y_1} \subset \mathcal{O}_{X_1}$ . Each component of  $Y_2 = \pi_2^{-1}(Y_1)$  is of codim 1.

Let  $X :=$  the normalization of  $X_2$  and  $\pi: X \rightarrow \text{Spec}(R)$  be the induced map. The morphism  $\phi: X \rightarrow X_2$  is finite since  $X_2 \rightarrow \text{Spec} R$  is of finite type, and  $R$  is a Nagata ring (see [M]). Let  $Y = \pi^{-1}(\mathfrak{m})_{\text{red}}$  be the reduced closed subscheme of  $X$ . Then  $\dim(Y) = \dim(Y_1)$  and each component of  $Y$  is of codimension 1 in  $X$ . Consider the finite subset of  $Y$  defined as

$$E = \{\{Y_{\text{sing}} \cup X_{\text{sing}}\} \cap \{\text{height 1 primes of } Y\}\} \cup \{\text{minimal primes of } Y\}.$$

We remark here that  $\pi: X \rightarrow \text{Spec } R$  is a birational surjective morphism of integral schemes. For any  $\mathfrak{p} \in V_R$ , the local ring  $R_{\mathfrak{p}}$  is either the quotient field of  $R$ , or a discrete valuation ring, which is a maximal proper subring of the quotient field. Hence the set  $\pi^{-1}(\mathfrak{p}) = \{x\}$  consists of a single point, and  $\mathcal{O}_{x,X} \cong R_{\mathfrak{p}}$ .

Let  $\mathcal{E} = \phi^* \pi_2^* \mathcal{M}$ . Then  $\mathcal{E}$  is a locally free sheaf on  $X$  of rank  $r$  generated by  $\tilde{g}_1, \dots, \tilde{g}_N$ , where  $\tilde{g}_i = \phi^* \pi_2^*(g_i)$ . Let  $\tilde{\psi}_i: \mathcal{O}_X \rightarrow \mathcal{E}$  be the  $\mathcal{O}_X$ -linear maps given by

$$1 \mapsto \alpha_{i1} \tilde{g}_1 + \dots + \alpha_{iN} \tilde{g}_N \quad \text{for } 1 \leq i \leq r,$$

where  $\alpha = (\alpha_{ij}) \in k^{rN}$ . Let  $\tilde{\psi}$  be the map  $\tilde{\psi}_1 \wedge \dots \wedge \tilde{\psi}_r: \mathcal{O}_X \rightarrow \wedge^r \mathcal{E}$ . Consider the ideal sheaf

$$\mathcal{T}_\alpha = \tilde{\psi}^*((\wedge^r \mathcal{E})^*) \subset \mathcal{O}_X,$$

where  $\tilde{\psi}^*: (\wedge^r \mathcal{E})^* \rightarrow \mathcal{O}_X$  is the dual of  $\tilde{\psi}$ . If for some  $x \in X$  the  $\mathcal{O}_{x,X}$ -module  $\mathcal{E}_x$ , which is free of rank  $r$ , is generated by, say,  $g_1, \dots, g_r$ , and if  $g_k = a_{1k}g_1 + \dots + a_{rk}g_r$ , where  $k \geq r+1$  and  $a_{ik} \in \mathcal{O}_{x,X}$ , then  $\mathcal{T}_{\alpha x} (\subset \mathcal{O}_{x,X})$  is generated by the element  $\det([\alpha_{ij} + \sum_{k \geq r+1} \alpha_{ik} a_{jk}],_{r \times r})$ . Note that  $\mathcal{T}_\alpha$  is the analogue for  $X$  and  $\mathcal{E}$  of the ideal  $T_\alpha$  defined earlier for the  $R$ -module  $M$ ; further, for  $\mathfrak{p} \in V_R$ , we have  $\mathcal{O}_{\mathfrak{p},X} \cong R_{\mathfrak{p}}$ , and  $\mathcal{T}_{\alpha \mathfrak{p}} = T_{\alpha \mathfrak{p}}$ .

Similarly, using the  $\mathcal{O}_Y$ -module  $i^* \mathcal{E}$ , we can define an ideal sheaf  $\tilde{\mathcal{T}}_\alpha \subset \mathcal{O}_Y$  for each  $\alpha \in k^{rN}$ ; then  $\tilde{\mathcal{T}}_\alpha$  equals the image of  $\mathcal{T}_\alpha$  under  $\mathcal{O}_X \rightarrow \mathcal{O}_Y$ .

*Reduction* Let  $D_\alpha$  be the closed subscheme of  $X$  with ideal sheaf  $\mathcal{T}_\alpha$ , and let  $\bar{D}_\alpha$  be the closed subscheme of  $Y$  with ideal sheaf  $\tilde{\mathcal{T}}_\alpha$ . Since for  $\mathfrak{p} \in V_R$ , we have  $\mathcal{O}_{\mathfrak{p},X} \cong R_{\mathfrak{p}}$ , it is enough to prove that for general  $\alpha \in k^{rN}$ , and for  $\mathfrak{p} \in V_R$ , either  $\mathcal{T}_{\alpha, \mathfrak{p}} = \mathcal{O}_{\mathfrak{p},X}$  or  $\mathcal{T}_{\alpha, \mathfrak{p}}$  is the maximal ideal of  $\mathcal{O}_{X, \mathfrak{p}}$ .

n) for  $y \in Y$  with  $\dim \mathcal{O}_{y,Y} = 1$ , either  $\mathcal{T}_{\alpha,y} = \mathcal{O}_{y,Y}$  or  $\mathcal{T}_{\alpha,y}$  is the maximal ideal.

*Proof of the claim.* Assume that for general  $\alpha$ , the statements i) and ii) hold. Fix such an  $\alpha$ . For  $x \in \text{reg}(X)$  of codimension 1, the subscheme  $\{x\} \cap Y$  of  $Y$  is of codimension  $\leq 1$ . Therefore there exists a  $y \in Y$  such that codimension of  $y$  in  $Y$  is  $\leq 1$  and  $y \in \{x\}$ . Since  $Y$  is of pure codimension 1 in  $X$ , the codimension of  $y$  in  $X$  is  $\leq 2$ . Moreover, if  $x \in Y$  then  $x \in E$ , hence one can choose  $y = x$  in that case.

a) If  $\mathcal{T}_{\alpha,x} = \mathcal{O}_{x,X}$ , then there is nothing to prove.

b) If  $y \in E$  then  $\bar{\mathcal{T}}_{\alpha,y} = \mathcal{O}_{y,Y}$ , which implies  $\mathcal{T}_{\alpha,x} = \mathcal{O}_{x,X}$ .

c) If  $y \notin E$  and  $\bar{\mathcal{T}}_{\alpha,x} \neq \mathcal{O}_{x,X}$  then by ii),  $\dim \mathcal{O}_{y,Y} = 1$  and  $\bar{\mathcal{T}}_{\alpha,y}$  is the maximal ideal of  $\mathcal{O}_{y,Y}$ . Since  $\mathcal{O}_{y,Y}$  is a regular local ring of dim 1 and  $\mathcal{O}_{y,X}$  is a regular local ring of dim 2, the generator of  $\mathcal{T}_{\alpha,y}$  is a part of a regular system of parameters of  $\mathcal{O}_{y,X}$ . Therefore  $\mathcal{T}_{\alpha,x}$  is a prime ideal of height 1 contained in  $\mathcal{O}_{x,X}$ . Hence  $\mathcal{T}_{\alpha,x}$  is the maximal ideal of  $\mathcal{O}_{x,X}$ .

*Proof of i) and ii) of the claim.* The sheaf  $\mathcal{E}|_Y$  is a locally free  $\mathcal{O}_Y$ -module of rank  $r$ . Therefore by lemma 8, we see that for the given finite set  $E$  of  $Y$ , the condition i) of the claim holds for general  $\alpha$ .

To prove the part ii) of the claim, it is enough to prove that, for all codimension 1 points  $y \in Y_{\text{reg}}$  and for general  $\alpha$ , the ring  $\mathcal{O}_{y,Y}/\mathcal{T}_y$  is regular.

Consider the separable extension  $k \subset K$ , where  $K$  is the algebraic closure of  $k$ . As  $Y$  is reduced and of finite type over  $k$ , the scheme  $Y \times_k \text{Spec}(K)$  is a reduced scheme of finite type over  $K$ . This also implies that the canonical map  $P: Y \times_k K \rightarrow Y$  is a faithful, affine map all of whose fibres are reduced. Moreover the canonical map  $P_\alpha: \mathcal{O}_{D_\alpha} \rightarrow \mathcal{O}_{D_\alpha \times K}$  is a faithful extension. Therefore, it is enough to prove the result for  $Y \times_k K$ , and in fact for the irreducible components of  $Y \times_k K$ . In other words we can assume  $k$  is algebraically closed and  $Y$  is an irreducible variety over  $k$ .

Consider the variety  $A_k^{rN} \times_k Y$  and the closed subset

$$C_0 = \{(\alpha_{ij}, \mathbf{p}) | \mu(v_1 \wedge \cdots \wedge v_r) = 0 \text{ in } (\wedge^r \mathcal{E})|_Y \otimes_{\mathcal{O}_{X_0}} k(\alpha_{ij}, \mathbf{p})\}$$

where  $v_i = X_{i1}g_1 + \cdots + X_{iN}g_N$  for  $k^{rN} = \text{Spec} k[X_{ij}]$  and where  $\mu$  is induced by  $X_{ij} \mapsto \alpha_{ij}$ . For the canonical map  $P_2: (C_0)_{\text{reg}} + A_k^{rN}$  there exists a nonempty open set  $V \subseteq A_k^{rN}$  such that  $P_2: P_2^{-1}(V) \rightarrow V$  is a smooth map (see [Ha], III, Corollary 10.7). Therefore (see [Ha], III, Corollary 10.2) for each  $\alpha \in V$ , the ring  $\mathcal{O}_{(C_0)_{\text{reg}}} \otimes_{\mathcal{O}_{X_0}} k(\alpha)$  is regular, i.e. for each  $\mathbf{p} \in Y$  in the fibre over  $\alpha$ , the ring  $\mathcal{O}_{\mathbf{p},Y}/\bar{\mathcal{T}}_{\alpha,\mathbf{p}}$  is regular.

Now it remains to prove that, for general  $\alpha$ , the subvariety  $(C_0)_{\text{sing}} \cap P_2^{-1}(\alpha)$  is of codimension  $> 1$  in  $Y_{\text{reg}}$ . But  $Y_{\text{reg}}$  is covered by finitely many open affine varieties  $\text{Spec}(R_i)$  such that  $\Gamma(\mathcal{E}_Y|_{R_i})$  is a free  $R_i$ -module of rank  $r$  with a basis given by a subset of cardinality  $r$  of the given generators  $g_1, \dots, g_N$ . Then  $A_k^{rN} \times Y$  has the corresponding open cover  $\{K_i = A_k^{rN} \times \text{Spec}(R_i)\}_i$ . If for a given such  $R_{i_0}$ , the submodule  $\Gamma(\mathcal{E}_Y|_{\text{Spec}(R_{i_0})})$  is generated by say  $g_1, \dots, g_r$  and  $g_k = a_{1k}g_1 + \cdots + a_{rk}g_r$  for  $k \geq r+1$ , where  $a_{ij} \in R_{i_0}$ , then the coordinate ring of

$$C_0|_{K_{i_0}} \text{ is } S_0 = k[X_{ij}] \otimes_k R_{i_0} \left/ \left( \det \left( X_{ij} + \sum_{k=r+1}^N a_{jk} X_{jk} \right)_{r \times r} \right) \right.$$

By changing variables we can assume that  $S_0 = k[\{Z_{ik}\}_{r \times N}] \otimes R_{i0}/(\det(Z_{ij})_{r \times r})$ . Therefore the coordinate ring of

$$(C_0|_{K_{i0}})_{\text{sing}} = (S_0)_{\text{sing}} = (k[Z_{ij}]/(\det(Z_{ij})))_{\text{sing}} \otimes_k R_{i0}.$$

But

$$\text{Spec}(k[Z_{ik}]/(\det(Z_{ij})))_{\text{sing}}$$

is a subscheme of codimension  $\geq 2$  in  $\text{Spec}[Z_{ik}]$ , since  $k[Z_{ik}]/(\det(Z_{ij}))$  is a normal domain (this is a particular case of [EH], Corollary (4)). Therefore  $(C_0)_{\text{sing}}|_{K_{i0}}$  is of codimension  $\geq 2$  in  $K_{i0}$ , which implies that  $(C_0|_{K_{i0}})_{\text{sing}} \cap P_2^{-1}(\alpha)$  is of codimension  $\geq 2$  in  $R_{i0}$ , for general  $\alpha \in k^{rN}$ . Therefore  $(C_0)_{\text{sing}} \cap P_2^{-1}(\alpha)$  is of codimension  $\geq 2$  in  $Y_{\text{reg}}$ .  $\square$

*Proof of theorem 4.* By lemma 8, it is enough to prove the existence of such a non-empty open set for all primes in  $V_R \setminus F$ , where  $F$  is some finite subset of  $V_R$ .

Now we will consider below a birational projective morphism

$$\ell: X \rightarrow \text{Spec}(R)$$

such that  $\ell^{-1}(U) \cong U$  for some open set  $U \subset \text{Spec } R$  containing  $(V_R \setminus F)$  for some finite subset  $F$  of  $V_R$ . Then we will consider a locally free sheaf  $\mathcal{H}$  on  $X$  such that  $\mathcal{H}|_{\ell^{-1}(U)} \cong \tilde{M}|_U$ , where  $\tilde{M}$  is the coherent sheaf on  $\text{Spec}(R)$  corresponding to  $M$ . Therefore it will be enough to prove the assertion for  $\mathcal{O}_X$  and  $\mathcal{H}$  instead of  $R$  and  $M$ .

Let  $X_1 = \text{Proj}(\text{Sym}_R(\wedge^r M)/(\text{torsion}))$  be the Nash blow up, which is an integral scheme, and let  $g_1: X_1 \rightarrow \text{Spec}(R)$  be the canonical map. Then  $g_1^{-1}(U_R) \cong U_R$  (see remark (1) of section 2). By lemma 2, the sheaf of  $\mathcal{O}_{X_1}$ -modules  $\mathcal{E} = (g_1^* \tilde{M})/(\text{torsion})$  is locally free of rank  $r$ , generated by the set  $\{g_i \otimes 1 \mid 1 \leq i \leq N\}$ , where  $\{g_1, \dots, g_N\}$  is a set of generators for  $M$ . Therefore, by the universal property of the Grassmannian (see [E.G.A.], section 9.7), for a given set of generators  $\{g_1, \dots, g_N\}$  of  $M$  there exists a morphism  $f_1: X_1 \rightarrow \text{Gr}_k(r, N)$  such that the diagram

$$\begin{array}{ccc} k^N \otimes_k \mathcal{O}_{X_1} & \xrightarrow{\quad} & \mathcal{E} \\ & \searrow & \nearrow \cong \tilde{f}_1 \\ & f_1^* \Omega & \end{array}$$

is commutative, where  $\Omega$  denotes the universal quotient bundle on  $\text{Gr}_k(r, N)$ . Henceforth, for the sake of brevity, we denote  $\text{Gr}_k(r, N)$  by  $\text{Gr}$ .

Let  $D_k(X_1)$  be the sheaf of finite differentials on  $X_1$ , which is a coherent  $\mathcal{O}_{X_1}$ -module (see [K] page 197). For any affine open subset  $U = \text{Spec}(A)$  of  $X_1$ , the  $A$ -module  $D_k(X_1)(U) = D_k(A)$ . By the universal property of the Kähler differentials we have an  $\mathcal{O}_{X_1}$ -linear surjective map  $\gamma: \Omega_{X_1/k}^1 \rightarrow D_k(X_1)$ , where  $\Omega_{X_1/k}^1$  is the sheaf of Kähler differentials of  $X_1$  relative to  $k$  (see [Ha] page 175, for example). Therefore there exists an  $\mathcal{O}_{X_1}$ -linear map  $\psi: f_1^* \Omega_{\text{Gr}/k}^1 \rightarrow D_k(X_1)$  which factors through  $\gamma$ .

Now we prove that there exists a set of generators  $\{g_1, \dots, g_N\}$  of  $M$ , as described in statement of the proposition, such that the map  $\psi: f_1^* \Omega_{\text{Gr}/k}^1 \rightarrow D_k(X_1)$  is not zero at any stalk.

*Lemma 11.* With the notations as above, we can choose generators  $\{g_1, \dots, g_N\}$  for  $M$



with the following properties. If  $\{m_1, \dots, m_k\}$  is a given set of generators of  $M$ , and if  $\{g_1, \dots, g_N\} \supset \{m_1, \dots, m_k, m\}$  for a suitable  $m \in M$ , then the map  $\psi: f_1^* \Omega_{\text{Gr}/k}^1 \rightarrow D_k(X_1)$  is not zero at any stalk. If  $\{x_1, \dots, x_n\}$  is a set of generators for the maximal ideal  $\mathfrak{m}$  of  $R$ , then we may take  $m = x_i m_j$  for suitable  $i, j$ .

*Proof.* First of all, for any given set of generators  $\{g_1, \dots, g_N\}$  of  $M$  containing  $\{m_1, \dots, m_k\}$ , we explicitly write the map

$$\psi: f_1^* \Omega_{\text{Gr}/k}^1 \rightarrow D_k(X_1).$$

Let  $T$  run over the subsets of  $\{1, 2, \dots, k\}$  of cardinality  $r$ , and for  $T = \{i_1, \dots, i_r\}$  let  $X_T$  denote the open subset of  $X_1$  on which the  $\mathcal{O}_{X_1}$ -module  $f_1^* \tilde{M}$  is globally generated by  $\{m_{i_1}, \dots, m_{i_r}\}$ . Then  $X_1 = \cup X_T$ . Let  $\{q_1, \dots, q_N\}$  be the global sections of  $\mathfrak{Q}$  such that  $f_1(q_i) = g_i$  for  $1 \leq i \leq N$ . For  $T = \{i_1, \dots, i_r\}$ , let  $U_T$  denote the open subset of  $\text{Gr}$  on which  $\{q_{i_1}, \dots, q_{i_r}\}$  generates  $\mathfrak{Q}$ . Then  $f_1^{-1}(U_T) = X_T$ . Since  $\{f_1^{-1}(U_T)\}_T$  is an open cover of  $X_1$  and  $\{U_T\}_T$  is an affine open cover of  $f_1(X_1) \subset \text{Gr}$ , to describe the map  $\psi$  it is enough to describe the maps

$$\psi|_{f_1^{-1}(U_T)}: \Omega_{\text{Gr}/k}^1 \otimes_{\mathcal{O}_{\text{Gr}}} \mathcal{O}_{f_1^{-1}(U_T)} \rightarrow D_k(f_1^{-1}(U_T)).$$

Without loss of generality, we may assume that  $T = \{1, 2, \dots, r\}$ . Now

$$U_T \cong \mathbb{A}^{r(N-r)} = \text{Spec } k[X_{11}, \dots, X_{1(N-r)}, \dots, X_{r1}, \dots, X_{r(N-r)}]$$

and  $\mathcal{O}_{U_T} \rightarrow \mathcal{O}_{X_T}$  is given by  $X_{ij} \rightarrow b_{ij}$ , where  $g_{j+r} = \sum_{i=1}^r b_{ij} m_i$ .

Now to prove the assertion, it is enough to prove the existence of the generators  $\{g_1, \dots, g_N\}$  of  $M$ , as described in the above lemma, such that

$$\psi|_{\mathcal{O}_{f_1^{-1}(U_T)}} 1_L: \Omega_{\text{Gr}/k}^1 \otimes_{\mathcal{O}_{\text{Gr}}} L \rightarrow D_k(X_1) \otimes_{\mathcal{O}_{\text{Gr}}} L = D_k(L)$$

is nonzero, where  $L$  denotes the function field of  $f_1^{-1}(U_T)$  (which is same as the quotient field of  $R$ ). But  $D_k(X) \otimes_{\mathcal{O}_{X_1}} L = D_k(R) \otimes L$  and  $\text{rank } D_k(R) \geq \dim(R) > 0$  (see [K], section 13). Therefore, as  $D_k(R)$  is an  $R$ -module generated by the images  $dx_i$  of a set of generators  $\{x_1, \dots, x_n\}$  for the maximal ideal  $\mathfrak{m}$  of  $R$ , we can find  $x_i$  with  $dx_i \neq 0$  in  $D_k(X_1) \otimes L$ .

Let us assume without loss of generality that the set  $\{m_1, \dots, m_r\}$  generates  $M$  on a nonempty open set  $X_T \cap U_R$  of  $X_1$ . Now let

$$\{g_1, \dots, g_N\} \supset \{m_1, \dots, m_r, x_i m_1\}, \quad g_N = x_i m_1 = x_i g_1,$$

where  $dx_i \neq 0$ , as above. For  $T = \{1, 2, \dots, r\}$  the map  $\mathcal{O}_{U_T} \rightarrow \mathcal{O}_{X_T}$  takes  $X_{1(N-r)}$  to  $x_i$ ; therefore  $\psi(d(X_{1(N-r)})) = dx_i$ . Hence  $\psi \otimes_{\mathcal{O}_{f_1^{-1}(U_T)}} 1_L \neq 0$ .  $\square$

Now, continuing the proof of the proposition, we define

$$\mathcal{F} = \psi(f_1^* \Omega_{\text{Gr}/k}^1).$$

Therefore  $\mathcal{F}$  is a nonzero sheaf of  $\mathcal{O}_{X_1}$ -modules which is of positive rank, say  $m$ . Let  $X_2$  be the Nash blow up of  $X_1$  associated to  $\mathcal{F}$ , i.e.,

$$X_2 = \text{Proj}(\text{Sym}_{\sigma_*}(\wedge^m \mathcal{F})/(\text{torsion}))$$

open subset  $V$  of  $U_R$ . Let

$$X = \text{Proj}(\oplus_{m \geq 0} \mathcal{I}_{Y_2}^m)$$

be the blow-up of  $X_2$  along  $Y_2$ , where  $\mathcal{I}_{Y_2}$  is the ideal sheaf corresponding to the reduced closed subscheme  $Y_2 := (g_2 \circ g_1)^{-1}(\mathbf{m})$ . If  $g: X \rightarrow X_1$  is the canonical map, then  $g^*\mathcal{E}$  is a locally free  $\mathcal{O}_X$ -module; we will take  $\mathcal{H} = g^*\mathcal{E}$ . Let  $\ell: X \rightarrow \text{Spec}(R)$  be the canonical map.

There exists a canonical map  $h: X \rightarrow \text{Gr}$  such that the following diagrams are commutative.

$$\begin{array}{ccc} X & \xrightarrow{h} & \text{Gr} \\ g_3 \downarrow & & \uparrow f_1 \\ X_2 & \xrightarrow{g_2} & X_1 \end{array} \quad \begin{array}{ccc} k^N \otimes \mathcal{O}_X & \longrightarrow & g^*\mathcal{E} \\ & \searrow & \nearrow \cong \tilde{h} \\ & h^*\mathcal{Q} & \end{array}$$

By lemma 2, the  $\mathcal{O}_{X_2}$ -module  $(g_2^*\mathcal{F}/(\text{torsion}))$  is locally free of rank  $m$ ; therefore if  $g_3: X \rightarrow X_2$  is the canonical map, then the  $\mathcal{O}_X$ -sheaf  $\mathcal{G} := g_3^*(g_2^*\mathcal{F}/(\text{torsion}))$  is locally free of rank  $m$ , and by definition of  $\mathcal{F}$  there exists a surjection  $h^*\Omega_{\text{Gr}/k}^1 \rightarrow \mathcal{G}$ .

The closed subscheme  $Y := g_3^{-1}(Y_2)$  of  $X$  has two special properties: (i) it is a projective scheme over  $k$  and (ii)  $Y$  is an effective Cartier divisor in  $X$ . We will define the sets of "bad points", namely  $C(X)$  and  $C(Y)$ , which map to  $k^{rN}$  and satisfy the property that  $C(X) \cap Y = C(Y)$ , such that for  $\alpha \in k^{rN}$ , the fibre  $C(X)_\alpha$  of  $C(X)$  restricted to the open set  $V \subset U_R$  is the set of point of  $V$  on which the ring  $R/T_\alpha$  is not regular. Now because of property (i) it will be easier to prove that for general  $\alpha \in k^{rN}$ , the fibre of  $C(Y)$  is of codim.1, and then property (ii) will verify the same assertion for  $C(X)$ .

Let  $\mathbf{A}^{rN} = \text{Spec} k[Y_{11}, \dots, Y_{1N}, \dots, Y_{r1}, \dots, Y_{rN}]$  and let  $p_1: \mathbf{A}^{rN} \times \text{Gr} \rightarrow \mathbf{A}^{rN}$  and let  $p_2: \mathbf{A}^{rN} \times \text{Gr} \rightarrow \text{Gr}$  be the natural projection maps. Let  $C_0(\text{Gr})$  be the closed subscheme of  $\mathbf{A}^{rN} \times \text{Gr}$ , as defined before lemma 10, corresponding to the universal quotient bundle  $\mathcal{Q}$ . Let

$$\bar{d}: \mathcal{I}_{C_0(\text{Gr})}/\mathcal{I}_{C_0(\text{Gr})}^2 \rightarrow p_2^*(\Omega_{\text{Gr}/k}^1) \otimes_k \mathcal{O}_{C_0(\text{Gr})} \quad (7)$$

be the map defined in the lemma 10.

Similarly if  $p_1: \mathbf{A}^{rN} \times X \rightarrow \mathbf{A}^{rN}$  and  $p_2: \mathbf{A}^{rN} \times X \rightarrow X$  are the projections, then

$$\left( \sum_{j=1}^N g_j Y_{1j} \right) \wedge \left( \sum_{j=1}^N g_j Y_{2j} \right) \wedge \dots \wedge \left( \sum_{j=1}^N g_j Y_{rj} \right)$$

is a section of the invertible sheaf  $p_2^*(\wedge^k(g^*\mathcal{E}))$ . Let  $C_0(X) \subset \mathbf{A}^{rN} \times X$  denotes its zero scheme and let  $\mathcal{I}_{C_0(X)}$  be the corresponding ideal sheaf. Then  $\mathcal{I}_{C_0(X)} = \mathcal{I}_{C_0(\text{Gr})} \otimes \mathcal{O}_X$ ,  $\mathcal{O}_{C_0(X)} = \mathcal{O}_{C_0(\text{Gr})} \otimes \mathcal{O}_X$ , and we have the following  $\mathcal{O}_{C_0(X)}$  linear map

$$\mathcal{I}_{C_0(X)}/\mathcal{I}_{C_0(X)}^2 \xrightarrow{\theta} p_2^*(h^*\Omega_{\text{Gr}/k}^1) \otimes \mathcal{O}_{C_0(X)} \xrightarrow{\eta} p_2^*(\mathcal{G}) \otimes \mathcal{O}_{C_0(X)} \quad (8)$$

where the surjective map  $\eta$  is the pullback of the map  $h^*\Omega_{\text{Gr}/k}^1 \rightarrow \mathcal{G}$ , and where the  $\mathcal{O}_{C_0(X)}$ -linear map  $\theta$  is obtained by tensoring (7) with  $\mathcal{O}_X$ . Since the  $\mathcal{O}_{C_0(X)}$ -sheaf

$\mathcal{F}_{C_0(X)}/\mathcal{F}_{C_0(X)}^2$  is locally free of rank 1, there is a natural isomorphism

$$\mathcal{H}om_{\mathcal{O}_{C_0(X)}}(\mathcal{F}_{C_0(X)}/\mathcal{F}_{C_0(X)}^2, p_2^*(\mathcal{G}) \otimes \mathcal{O}_{C_0(X)}) \cong (\mathcal{F}_{C_0(X)}/\mathcal{F}_{C_0(X)}^2)^* \otimes (p_2^*(\mathcal{G}) \otimes \mathcal{O}_{C_0(X)}).$$

Let  $s$  be the section of  $(\mathcal{F}_{C_0(X)}/\mathcal{F}_{C_0(X)}^2)^* \otimes (p_2^*(\mathcal{G}) \otimes \mathcal{O}_{C_0(X)})$ , corresponding to the map  $\eta \circ \theta$  (see (8)). Then the following set defined as

$$C(X) = \{q \in \mathbf{A}^{rN} \times X \mid s \otimes k(q) = 0\}$$

is a closed subscheme of  $\mathbf{A}^{rN} \times X$ .

*Claim (1).* It is enough to prove that the general fibre of  $p_1: C(X) \rightarrow \mathbf{A}^{rN}$  is of codim.  $\geq 2$  in  $X$ .

*Proof of the claim.* Each of the following maps

$$X \xrightarrow{\pi_3} X_2 \xrightarrow{\pi_2} X_1 \xrightarrow{\pi_1} \text{Spec } R$$

is birational. Therefore there exists a maximal open set  $U \subseteq \text{Spec } R$  such that  $\ell^{-1}(U) \cong U$ , where  $\ell: X \rightarrow \text{Spec } R$  is the above composite map. Then  $U \supset V_R$ , since for each  $\mathbf{p} \in V_R$ , the local ring  $R_{\mathbf{p}}$  is either a field or a discrete valuation ring. Thus for each point  $\mathbf{p}$  of  $V_R$ , there exists a unique preimage  $\mathbf{q}$  is  $X$ , and an isomorphism of local rings  $\ell^*: R_{\mathbf{p}} \rightarrow \mathcal{O}_{X, \mathbf{q}}$ . We identify any element  $\mathbf{p}$  of  $V_R$  with its (unique) inverse images in  $X$  or  $X_i$ .

Now  $\mathcal{F}$  is a coherent subsheaf of  $D_k(X_1)$ ; since  $R$  is a complete local domain, the set of primes of  $V_R$ , for which either  $\mathcal{F}_{\mathbf{p}}$  or  $D_k(R)_{\mathbf{p}}$  is not free over  $R_{\mathbf{p}}$ , or  $\mathcal{F} \otimes k(\mathbf{p}) \rightarrow D_k(R) \otimes k(\mathbf{p})$  is not injective, is finite. Let  $F \subset V_R$  be this finite set. For  $\mathbf{p} \in V_R \setminus F$ , we have that  $\mathcal{F}_{\mathbf{p}} \rightarrow D_k(R)_{\mathbf{p}}$  is a split inclusion of finitely generated free  $R_{\mathbf{p}}$ -modules.

Suppose the general fibre of  $p_1: C(X) \rightarrow \mathbf{A}^{rN}$  is of codimension  $\geq 2$  in  $X$ . Then, for general  $\alpha = (\alpha_{ij}) \in k^{rN}$ , the fibre of  $C(X)$  over  $\alpha$  is disjoint from  $V_R \setminus F$ . Now for  $\mathbf{p} \in V_R \setminus F$ , the ideal  $\mathcal{F}_{C_0(X)} \otimes k(\alpha_{ij}) \otimes R_{\mathbf{p}} = T_{\alpha, \mathbf{p}}$ , (see (6) in the beginning of the section) which is a cyclic  $k(\alpha_{ij}) \otimes R_{\mathbf{p}}$ -module. Localising the equation (8) at  $\mathbf{p}$ , we get

$$T_{\alpha, \mathbf{p}}/T_{\alpha, \mathbf{p}}^2 \rightarrow p_2^*(\mathcal{G}) \otimes \mathcal{O}_{C_0(X)} \otimes k(\alpha_{ij}) \otimes R_{\mathbf{p}} \subset D_k(R)_{\mathbf{p}} \otimes (R_{\mathbf{p}}/T_{\alpha, \mathbf{p}}).$$

(The inclusion on the right is because  $\mathcal{G}_{\mathbf{p}} = \mathcal{F}_{\mathbf{p}} \subset D_k(R)_{\mathbf{p}}$  is a split inclusion of free  $R_{\mathbf{p}}$ -modules.) If  $\mathbf{p} \notin$  fibre of  $C(X)$  over  $\alpha$  then, by the definition of  $C(X)$ , the image of  $T_{\alpha, \mathbf{p}}/T_{\alpha, \mathbf{p}}^2$  contains a basic element of  $p_2^*(\mathcal{G}) \otimes \mathcal{O}_{C_0(X)} \otimes k(\alpha_{ij}) \otimes R_{\mathbf{p}}$ , hence of  $D_k(R)_{\mathbf{p}} \otimes (R_{\mathbf{p}}/T_{\alpha, \mathbf{p}})$ . The composite map  $T_{\alpha, \mathbf{p}}/T_{\alpha, \mathbf{p}}^2 \rightarrow D_k(R) \otimes R_{\mathbf{p}}/T_{\alpha, \mathbf{p}}$  is the natural map induced by the canonical derivation  $d: R_{\mathbf{p}} \rightarrow D_k(R)_{\mathbf{p}}$ . As  $R_{\mathbf{p}}$  is regular, by lemma 7 this implies that  $R_{\mathbf{p}}/T_{\alpha, \mathbf{p}}$  is regular, for every  $\mathbf{p} \in V_R \setminus F$ . Hence the claim.

*Reduction.* We can replace  $X$  by the closed subscheme  $Y := \ell^{-1}(\mathbf{m})$ .

Let  $i: Y \hookrightarrow X$  be the natural inclusion map. Then  $i^*(g^*\mathcal{G})$  is a locally free  $\mathcal{O}_Y$ -module of rank  $r$ . Let  $C_0(Y)$  and  $C(Y)$  be the closed subschemes of  $\mathbf{A}^{rN} \times Y$ , constructed in the same way as  $C_0(X)$  and  $C(X)$  are constructed for  $\mathbf{A}^{rN} \times X$ . Let  $\mathcal{F}_{C_0(Y)}$  be the ideal sheaf of  $C_0(Y)$ . It is easy to see from the definition of  $C_0(X)$  and  $C_0(Y)$  that  $C_0(X) \cap Y = C_0(Y)$  and  $\mathcal{F}_{C_0(Y)}$  is the image of  $\mathcal{F}_{C_0(X)}$  in  $\mathcal{O}_{\mathbf{A}^{rN} \times Y}$ , where  $\mathcal{O}_{\mathbf{A}^{rN} \times X} \rightarrow$

$\mathcal{C}_{\mathbf{A}^{rN} \times Y}$  is the canonical surjective map. Therefore the map  $\mathcal{O}_{C_0(X)} \rightarrow \mathcal{O}_{C_0(Y)}$  is surjective and  $\mathcal{J}_{C_0(Y)} = \mathcal{J}_{C_0(X)} \otimes \mathcal{O}_Y$ ,  $\mathcal{C}_{C_0(Y)} = \mathcal{C}_{C_0(X)} \otimes \mathcal{O}_Y$ . Now we have the following  $\mathcal{O}_{C_0(Y)}$ -linear map

$$\mathcal{J}_{C_0(Y)} / \mathcal{J}_{C_0(Y)}^2 \xrightarrow{\theta_1} p_2^*(i^* h^* \Omega_{\text{Gr}/k}^1) \otimes \mathcal{O}_{C_0(Y)} \xrightarrow{\eta_1} p_2^*(i^* \mathcal{G}) \otimes \mathcal{O}_{C_0(Y)} \quad (9)$$

which is same as the map obtained by tensoring map (8) with  $\mathcal{O}_Y$ ; therefore  $C(X) \cap Y = C(Y)$ . Hence  $(C(X) \otimes k(\alpha)) \cap Y = C(Y) \otimes k(\alpha)$ , for any  $\alpha \in \mathbf{A}^{rN}$ . This implies that

$$\text{codim}_Y(C(Y) \otimes k(\alpha)) \leq \text{codim}_X(C(X) \otimes k(\alpha)),$$

as  $Y$  contains all closed points of  $X$  and as the ideal sheaf  $\mathcal{J}_Y$  of  $Y$  is locally principal. Therefore it is enough to prove that the general fibre of  $C(Y)$  is of  $\text{codim.} \geq 2$  in  $Y$ .

Let  $Y_{\text{red}} = Y_1 \cup \dots \cup Y_l$ , where  $Y_i$  are the irreducible reduced components of  $Y$ . If for each  $Y_i$  we define  $C(Y_i)$  and  $C_0(Y_i)$ , as we have defined then for  $Y$ , then  $C_0(Y) \cap (Y_i \times \mathbf{A}^{rN}) = C_0(Y_i)$  and  $C(Y) \cap (Y_i \times \mathbf{A}^{rN}) = C(Y_i)$ . Therefore  $C(Y) = \cup C(Y_i)$  as sets, which implies that if the general fibre of  $C(Y_i)$  is of  $\text{codim.} \geq 2$  in  $Y_i$  for each  $i$ , then the general fibre of  $C(Y)$  is of  $\text{codim.} \geq 2$  in  $Y$ . Hence we can assume that  $Y$  is integral.

*Claim (2).* The fibre of  $C_0(Y)$  over  $t \in Y$ , i.e.,  $C_0(Y)_t$ , is an irreducible variety of dimension  $rN - 1$  for every  $t$ .

*Proof of the claim.* Since  $(\mathcal{O}_{\mathbf{A}^{rN}} \otimes \mathcal{O}_Y) \otimes \mathcal{J}_{C_0(\text{Gr})} = \mathcal{O}_Y \otimes_{\mathcal{O}_{\text{Gr}}} \mathcal{J}_{C_0(\text{Gr})} = \mathcal{J}_{C_0(Y)}$ , lemma 10 implies that the set  $C_0(Y)_t$ , where  $t \in Y$ , is described as

$$C_0(Y)_t = \{p \in \mathbf{A}_{k(t)}^{rN} \mid (\det(t_{ij}) \otimes k(t)) \otimes k(p) = 0\}.$$

But  $t_{ij}$  are part of a system of coordinates for  $\mathbf{A}_{k(t)}^{rN}$ ; therefore  $C_0(Y)_t$  is an irreducible variety of dimension  $rN - 1$  (see (EH), corollary (4)) for every  $t \in Y$ .

*Claim (3).* The fibre of  $C(Y)$  over  $t$ , i.e.,  $C(Y)_t$ , is a proper subset of  $C_0(Y)_t$ , for every  $t \in Y$ .

We assume the claim for the moment. Since  $\dim C_0(Y)_t = rN - 1$ , by the above claim we have  $\dim C(Y)_t \leq rN - 2$ , for every  $t \in Y$ . If  $C(Y) = C(Y)^1 \cup \dots \cup C(Y)^n$ , where  $C(Y)^j$  are the irreducible components of  $C(Y)$ , then  $\dim C(Y)_t^j \leq rN - 2$  for each  $1 \leq j \leq n$ . Therefore

$$\dim C(Y)^j \leq rN - 2 + \dim Y.$$

If the canonical projection  $\text{map } p_2: C(Y)^j \rightarrow \mathbf{A}^{rN}$  is not dominant, then the general fibre of  $C(Y)^j \rightarrow \mathbf{A}^{rN}$  is empty, hence has codimension  $\geq 2$ . If the projection map is dominant then

$$rN + \dim \text{ of general fibre of } C(Y)^j \text{ over } \mathbf{A}^{rN} = \dim C(Y)^j$$

*Proof of the claim (3).* For any  $t \in Y$ , the map

$$\mathcal{O}_{C_0(Y)} \otimes_{\mathcal{O}_Y} k(t) \rightarrow (\mathcal{I}_{C_0(Y)} / \mathcal{I}_{C_0(Y)}^2) \otimes k(t),$$

given by  $1 \mapsto \det(t_{ij})$ , is an isomorphism. Therefore to prove the claim, it is enough to prove that the composite map (obtained from the map (9))

$$\begin{aligned} (\mathcal{I}_{C_0(Y)} / \mathcal{I}_{C_0(Y)}^2) \otimes k(t) &\xrightarrow{\tilde{\theta}_1} p_2^*(i^* h^* \Omega_{\text{Gr}/k}^1) \otimes \mathcal{O}_{C_0(Y)} \otimes k(t) \\ &\downarrow \tilde{\eta}_1 \\ p_2^*(i^* \mathcal{G}) \otimes \mathcal{O}_{C_0(Y)} \otimes k(t) \end{aligned}$$

is not identically zero on  $C_0(Y)_t$ , as  $C(Y)_t$  is the set of points of  $C_0(Y)_t$  for which this composition is zero. The map  $\tilde{\theta}_1$  can be rewritten as

$$(\mathcal{I}_{C_0(\text{Gr})} / \mathcal{I}_{C_0(\text{Gr})}^2) \otimes_{\mathcal{O}_{\text{Gr}}} k(g) \otimes_{k(g)} k(t) \xrightarrow{\tilde{\theta}} (p_2^* \Omega_{\text{Gr}/k}^1) \otimes \mathcal{O}_{C_0(\text{Gr})} \otimes_{\mathcal{O}_{\text{Gr}}} k(g) \otimes_{k(g)} k(t)$$

and  $\tilde{\eta}_1$  can be rewritten as  $\mathcal{O}_{C_0(\text{Gr})} \otimes k(t)$ -linear map

$$(p_2^* \Omega_{\text{Gr}/k}^1) \otimes \mathcal{O}_{C_0(\text{Gr})} \otimes_{\mathcal{O}_{\text{Gr}}} k(t) \rightarrow \mathcal{G} \otimes_{\mathcal{O}_X} k(t) \otimes_{\mathcal{O}_{\text{Gr}}} \mathcal{O}_{C_0(\text{Gr})},$$

where  $g = (h \circ i)(t) \in \text{Gr}$  for the canonical map  $h \circ i: Y \rightarrow \text{Gr}$ . The map  $\tilde{\theta}_1 \equiv \bar{d} \otimes 1_{k(g)} \otimes 1_{k(t)}$ .

Since  $\Omega_{\text{Gr}/k}^1$  and  $\mathcal{G}$  are locally free  $\mathcal{O}_{\text{Gr}}$  and  $\mathcal{O}_X$  modules of ranks  $r(N-r)$  and  $m$  respectively, we observe that

$$(\Omega_{\text{Gr}/k}^1) \otimes_{\mathcal{O}_{\text{Gr}}} \mathcal{O}_{C_0(\text{Gr})} \otimes_{\mathcal{O}_{\text{Gr}}} k(g) \otimes_{k(g)} k(t) \cong B^{r(N-r)}$$

and

$$(\mathcal{G} \otimes_{\mathcal{O}_X} k(t)) \otimes_{k(g)} \mathcal{O}_{C_0(\text{Gr})} \otimes k(g) \cong B^m,$$

where  $B := \Gamma(\mathcal{O}_{C_0(\text{Gr})} \otimes k(g)) = k(t)[\{t_{ij}\}_{r \times r}, \{Z_{il}\}_{r \times (N-r)}] / (\det(t_{ij}))$ . Now, by lemma 10, we conclude that the Image  $(\tilde{\theta}_1)$  is not contained in  $\ker(\tilde{\eta}_1)$ . Hence the claim.

## 6. A Bertini theorem for the general linked ideal

We give the definition of geometric linkage as given in [PS] for the affine case. For further properties of this notion, one can refer to it.

### DEFINITION

Let  $R/I$  and  $R/J$  be two quotients of a regular local ring  $R$ . Then  $R/I$  and  $R/J$  are said to be *geometrically linked* if

- (i)  $R/I$  and  $R/J$  are equidimensional, without embedded components and without common irreducible components.
- (ii)  $R/(I \cap J)$  is a complete intersection in  $R$ .

Now we recall *Serre's conditions*  $(R_s)$  and  $(S_r)$ .

- (i) A ring  $R$  is said to satisfy condition  $(R_s)$  if the ring  $R_{\mathfrak{p}}$  is regular for all  $\mathfrak{p} \in \text{Spec } R$  of height  $\leq s$ ,

(ii) A ring  $R$  is said to satisfy condition  $(S_r)$ , if for all  $\mathfrak{p} \in \text{Spec } R$ ,  $\text{depth}(R_{\mathfrak{p}}) \geq \min(\text{ht } \mathfrak{p}, r)$

**Theorem 5.** Let  $R$  be a local ring containing an infinite perfect field  $k$  with  $\hat{R}$  equidimensional. Let  $I$  be an ideal of height  $r$ .

(i) If  $R$  satisfies  $(S_r)$  then for any given set of generators  $\{a_1, \dots, a_k\}$  of  $I$ , the ideal

$$\mathfrak{F}_{\alpha} = \left\langle \sum_j \alpha_{1j} a_j, \dots, \sum_j \alpha_{rj} a_j \right\rangle$$

is generated by an  $R$ -sequence, for general  $\alpha = (\alpha_{ij}) \in k^{r \times k}$ .

There exists a set of generators of  $I$ , say  $g_1, \dots, g_N$ , with the following properties.

(ii) If  $R$  is regular and  $I$  is equidimensional of height  $r$  (i.e., every minimal prime ideal of  $I$  has height  $r$ ) and is a reduced ideal then the ideal  $J_{\alpha} := (\mathfrak{F}_{\alpha} : I)$  is reduced and is geometrically linked to  $I$  via  $\mathfrak{F}_{\alpha}$ , for general  $\alpha \in k^{r \times N}$ .

(iii) If  $R$  is a regular local Nagata ring and  $I$  is an ideal as in (ii) above which is a local complete intersection in  $\text{codim.} 1$  (i.e.,  $I_{\mathfrak{p}}$  is generated by an  $R_{\mathfrak{p}}$ -sequence for  $\text{ht}(r+1)$  prime ideals  $\mathfrak{p}$  in  $R$  which contain  $I$ ) then for general  $\alpha \in k^{r \times N}$ , (a)  $R/J_{\alpha}$  satisfies Serre's condition  $(R_1)$  and is regular on  $D(I)$ , and (b) if  $R/I$  is seminormal then  $R/\mathfrak{F}_{\alpha}$  is seminormal.

*Proof.*

(i) Suppose  $R$  satisfies Serre's condition  $(S_r)$ .

By proposition 1, the ideal  $\mathfrak{F}_{\alpha} = \langle \sum_j \alpha_{1j} a_j, \dots, \sum_j \alpha_{rj} a_j \rangle$  is of height  $r$  for general  $\alpha \in k^{r \times N}$ . Since  $R$  satisfies  $(S_r)$ , this implies that  $\mathfrak{F}_{\alpha}$  is generated by an  $R$ -regular sequence.

(ii) Suppose  $R$  is regular and  $I$  is an equidimensional radical ideal of height  $r$ .

Let  $\{a_1, \dots, a_k\}$  be a set of generators of  $I$  and let  $\{g_1, \dots, g_N\}$  be the set  $\{a_1, \dots, a_k, x_1 a_1, x_1 a_2, \dots, x_d a_k\}$  where  $\{x_1, \dots, x_d\}$  is a some set of generators of  $\mathfrak{m}$ . By Theorem 2, for general  $\alpha \in k^{r \times N}$  the ring  $R/\mathfrak{F}_{\alpha}$  is regular on  $D(I)$ . We fix a general  $\alpha \in k^{r \times N}$  for which  $\mathfrak{F}$  is generated by a regular sequence and  $R/\mathfrak{F}_{\alpha}$  is regular on  $D(I)$ . Now we have

*Claim I.*  $\mathfrak{F}_{\alpha} = I \cap J_{\alpha}$ , where  $J_{\alpha}$  is a radical equidimensional ideal of height  $r$ , all of whose minimal primes are distinct from the minimal primes of  $I$ .

*Proof of the claim.* Since  $R$  satisfies  $(S_{r+1})$  and  $\mathfrak{F}$  is generated by a regular sequence, the ring  $R/\mathfrak{F}_{\alpha}$  satisfies condition  $(S_1)$ ; which implies

$$\{\text{Ass primes}\} = \{\text{min. primes}\} = \{\mathfrak{p} \in \text{Spec}(R/\mathfrak{F}_{\alpha}) \text{ which are of ht } r \text{ in } R\}.$$

Since  $I$  is radical,  $\{\text{Ass primes of } R/I\} = \{\text{min. primes}\}$ , and by hypothesis each of them is of height  $r$ . Therefore

$$\{\text{Ass primes of } R/I\} \subseteq \{\text{Ass primes of } R/\mathfrak{F}_{\alpha}\}.$$

Hence the primary decomposition of  $\mathfrak{F}_{\alpha}$  is given by

$$\mathfrak{F}_{\alpha} = Q_1 \cap \dots \cap Q_{r_1} \cap Q'_1 \cap \dots \cap Q'_{r_2},$$

$Q'_i$  corresponds to some prime of  $D(I)$ . Since  $R/I$  is a reduced ring, it is regular at every minimal prime  $\mathfrak{p}$ , and hence the ideal  $I_{\mathfrak{p}}$  is generated by some  $r$  elements of the set  $\{g_1, \dots, g_N\}$ . Therefore, by lemma 8,  $I_{\mathfrak{p}} = \mathfrak{F}_{\alpha\mathfrak{p}}$  for all minimal primes of  $R/I$ . Hence  $Q_i$  are all prime and  $Q_1 \cap \dots \cap Q_{r_1} = I$ . By theorem 2, for general  $\alpha$ , the ring  $(R/\mathfrak{F}_{\alpha})_{\mathfrak{p}}$  is regular for all  $\mathfrak{p} \in D(I)$ . This implies that all  $Q'_i$  are also primes. Taking  $J_{\alpha} = Q'_1 \cap \dots \cap Q'_{r_2}$ , we have proved the claim.

*Claim II.*  $I = (\mathfrak{F}_{\alpha} : J_{\alpha})$  and  $J_{\alpha} = (\mathfrak{F}_{\alpha} : I)$

*Proof of the claim.* Since  $\mathfrak{F}_{\alpha} = I \cap J_{\alpha} \supseteq IJ_{\alpha}$ , we have  $I \subseteq (\mathfrak{F}_{\alpha} : J_{\alpha})$ . Moreover, for  $\mathfrak{p} \in \text{Ass}(I)$ , we have  $I_{\mathfrak{p}} = (\mathfrak{F}_{\alpha\mathfrak{p}} : J_{\alpha\mathfrak{p}})$ , as  $I_{\mathfrak{p}} = \mathfrak{F}_{\alpha\mathfrak{p}}$  and  $J_{\alpha\mathfrak{p}} = R_{\mathfrak{p}}$ . Therefore primary decomposition of  $(\mathfrak{F}_{\alpha} : J_{\alpha})$  is given by

$$(\mathfrak{F}_{\alpha} : J_{\alpha}) = \tilde{P}_1 \cap \dots \cap \tilde{P}_{s_1} \cap \tilde{Q}'_1 \cap \dots \cap \tilde{Q}'_{s_2},$$

where  $\tilde{P}_1 \cap \dots \cap \tilde{P}_{s_1} = I$ . This implies  $(\mathfrak{F}_{\alpha} : J_{\alpha}) \subseteq I$ . Hence  $(\mathfrak{F}_{\alpha} : J_{\alpha}) = I$ .

Similarly we can prove that  $J_{\alpha} = (\mathfrak{F}_{\alpha} : I)$ . In other words  $J_{\alpha}$  is geometrically linked to  $I$  via  $\mathfrak{F}_{\alpha}$ , for general  $\alpha$ .

(iii) Suppose  $R$  is a regular local Nagata ring and  $I$  is an ideal as in (ii) above which is a complete intersection in codim.1.

(a) Since on  $D(I)$ ,  $J_{\alpha} = \mathfrak{F}_{\alpha}$  and  $R/\mathfrak{F}_{\alpha}$  is regular for general  $\alpha$ , the ring  $R/J_{\alpha}$  is regular on  $D(I)$  for general  $\alpha$ . Hence it is enough to prove that  $R/J_{\alpha}$  is  $(R_1)$  for general  $\alpha$ .

Taking  $M = I/I^2$  and  $A = R/I$ , and applying Theorem 4, we get that for general  $\alpha$ , the ring  $R_{\mathfrak{p}}/(I_{\mathfrak{p}} + T_{\alpha\mathfrak{p}})$  is regular for all height 1 regular primes (i.e., where  $(R/I)_{\mathfrak{p}}$  is regular) of  $R/I$ . We recall that

$$T_{\alpha\mathfrak{p}} = (\phi_1^* \wedge \dots \wedge \phi_r^*)(\wedge^r M_{\mathfrak{p}}^*) \subset R_{\mathfrak{p}}$$

(See equation 6). In fact if  $I_{\mathfrak{p}}$  is generated by  $g_1, \dots, g_r$  and if  $g_k = a_{1k}g_1 + \dots + a_{rk}g_r$ , where  $k \geq r+1$  and  $a_{ik} \in R_{\mathfrak{p}}$ , then  $T_{\alpha\mathfrak{p}} = \det([\alpha_{ij} + \sum_k a_{jk}\alpha_{ik}]_{r \times r})R_{\mathfrak{p}}$ . Since height  $I = r$  and  $I$  is a local complete intersection in codim.1, the ideal  $I$  is generated by  $r$  elements, at prime ideals of height 1 in  $R/I$ . Since  $R/I$  is a reduced ring, the nonregular locus is of codim.  $\geq 1$ . Therefore the set  $F$ , of all minimal prime ideals and non regular prime ideals of height 1 (of  $R/I$ ), is finite. Hence, by lemma 8, for general  $\alpha$  and for such primes, the ideal  $T_{\alpha\mathfrak{p}} = R_{\mathfrak{p}}$ . Consider a height 1 prime ideal of  $R/J_{\alpha\mathfrak{p}}$ . If  $T_{\alpha\mathfrak{p}} = R_{\mathfrak{p}}$  then  $J_{\alpha\mathfrak{p}} = R_{\mathfrak{p}}$  as  $T_{\alpha\mathfrak{p}} \subseteq J_{\alpha\mathfrak{p}}$ . If  $I_{\mathfrak{p}} = R_{\mathfrak{p}}$  then  $J_{\alpha\mathfrak{p}} = \mathfrak{F}_{\alpha\mathfrak{p}}$  which implies  $(R/J_{\alpha})_{\mathfrak{p}}$  is regular. Hence from now onwards, we assume that

$$\mathfrak{p} \in \{V(I + J_{\alpha\mathfrak{p}})\} \cap \{(R/I)_{\text{Reg}}\} \cap \{\text{height 1 prime ideals of } (R/J_{\alpha})\},$$

where  $\alpha$  is a fixed element of the general set described above.

As remarked before, if  $I_{\mathfrak{p}}$  is generated by  $g_1, \dots, g_r$  and if  $g_k = a_{1k}g_1 + \dots + a_{rk}g_r$ , where  $k \geq r+1$  and  $a_{ik} \in R_{\mathfrak{p}}$ , then  $T_{\alpha\mathfrak{p}} = \det([\alpha_{ij} + \sum_k a_{jk}\alpha_{ik}]_{r \times r})$ . For convenience we denote  $\alpha_{ij} + \sum_k a_{jk}\alpha_{ik}$  by  $f_{ij}$ . Since  $\det([f_{ij}]) \neq \mathfrak{p}_{\mathfrak{p}}^2$  (as  $R_{\mathfrak{p}}/(I_{\mathfrak{p}} + T_{\alpha\mathfrak{p}})$  is regular), at least one of the size  $(r-1)$  minor of  $[f_{ij}]$  is invertible in  $R_{\mathfrak{p}}$ , by lemma 9. Let  $M_{ij}$  denote the matrix obtained by deleting  $i$ th row and  $j$ th column of  $[f_{ij}]$  and let  $R_{ij}$  denote the  $(i,j)$ th cofactor of  $[f_{ij}]$ . Without loss of generality we may assume that  $M_{rr}$  is

$$(\mathfrak{F}_\alpha)_p = \langle [f_{ij}][g_j]_{r \times 1} \rangle_p$$

$$= \left\langle \begin{bmatrix} 0 \\ M_{rr}^{-1} \\ \vdots \\ R_{1r} & \cdots & R_{rr} \end{bmatrix} [f_{ij}] \begin{bmatrix} g_1 \\ \vdots \\ g_r \end{bmatrix} \right\rangle_p$$

$$= \left\langle \begin{bmatrix} & & t_1 \\ & Id_{r-1} & \vdots \\ & & t_{r-1} \\ 0 & \cdots & \det(f_{ij}) \end{bmatrix} \begin{bmatrix} g_1 \\ \vdots \\ g_r \end{bmatrix} \right\rangle_p$$

$$= \langle g_1 + t_1 g_r, \dots, g_{r-1} + t_{r-1} g_r, (\det f_{ij}) g_r \rangle_p,$$

where  $t_i$  are some elements in  $R_p$ . Since  $R_p/(I_p + T_{ap})$  is regular and  $(I_p + T_{ap})$  is of height  $r+1$  in the regular local ring  $R_p$ , the generators  $g_1 + t_1 g_r, \dots, g_{r-1} + t_{r-1} g_r, g_r, \det(f_{ij})$  of  $I_p + T_{ap}$  forms a regular system of parameters of  $R_p$ . Therefore

$$I_{ap} := (g_1 + t_1 g_r, \dots, g_{r-1} + t_{r-1} g_r, \det(f_{ij}))_p$$

is a prime ideal of height  $r$  and  $R/I_{ap}$  is regular. This implies  $J_{ap} = I_{ap}$ , as  $J_{ap} \subseteq I_{ap}$  and  $J_{ap}$  is of height  $r$ . Therefore  $(R/J_\alpha)_p$  is regular. (b) Since  $(\mathfrak{F}_\alpha)_p = (g_1 + t_1 g_r, \dots, g_{r-1} + t_{r-1} g_r, \det(f_{ij}) g_r)_p$ , where  $g_1 + t_1 g_r, \dots, g_{r-1} + t_{r-1} g_r, g_r, \det(f_{ij})$  forms a regular system of parameters of  $R_p$ , the ring  $R_p/(\mathfrak{F}_\alpha)_p$  is seminormal (see [GT] theorem 8.1). Now since it is seminormal at all height 1 prime ideals and satisfies  $(S_2)$  (in fact is Cohen-Macaulay), by corollary 2.7 of [GT], it is seminormal.  $\square$

*Remark.* In the statement of the theorem, we have not specified the set of generators of the ideal  $I$ . The proof given shows that if  $\{a_1, \dots, a_k\}$  is any set of generators for  $I$ , and  $\{x_1, \dots, x_d\}$  is any set of generators of the maximal ideal of  $R$ , then the conclusions of the theorem are valid for the set  $S = \{a_1, \dots, a_k, x_1 a_1, \dots, x_i a_j, \dots, x_d a_k\}$  of generators of  $I$ . A small modification of the proof shows that it is enough to take any set of generators of  $I$  containing  $S$ .

Since in a regular local ring a geometric link of a Cohen-Macaulay ring is a Cohen-Macaulay ring (see [PS]) we have the

## COROLLARY 1

Let  $R$  be a regular local Nagata ring containing an infinite perfect field  $k$ ,  $I \subset R$  an equidimensional radical ideal of height  $r$ , such that

- (i)  $R/I$  is Cohen-Macaulay, and
- (ii)  $R/I$  is a local complete intersection in codimension 1.

Then, for the general  $\alpha \in k^{rN}$ , the linkage  $R/J_\alpha$  is normal and Cohen-Macaulay.

## Acknowledgements

I would like to thank my thesis advisor, Prof. R C Cowsik, for initiating me into the study of local Bertini theorems, and for suggesting this problem. In particular, he



suggested the approach to the proof of proposition 1 via associated graded rings; this led to the approach to the other results via blow ups. I thank Prof. Mohan Kumar for stimulating discussions on this subject. I express my deep gratitude to Dr V Srinivas for his many valuable suggestions, and for help in preparing this paper for publication. I am grateful to NBHM for financial support.

## References

- [EH] Eagon J A and Hochster M, Cohen-Macaulay rings, invariant theory and the generic perfection of determinantal loci, *Am. J. Math.* **93** (1971) 1020–1058
- [F] Flenner H, Die Sätze von Bertini für locale Ringe, *Math. Ann.* **229** (1977) 97–111
- [GT] Greco S and Traverso C, On seminormal schemes, *Compositio Math.* **40** (1980) 325–365
- [EGA] Grothendieck A and Dieudonné J, *Éléments de Géométrie Algébrique*, *Grundlehren Math. Band* 166 (Berlin: Springer-Verlag) (1971)
- [Ha] Hartshorne R, Algebraic geometry, *Grad. Texts in Math No. 52*, (New York: Springer-Verlag) (1977)
- [H2] Hartshorne R, Complete intersections and connectedness, *Am. J. Math.* **84** (1962) 497–508
- [HU] Hunecke C and Ulrich B, Generic Residual Intersections, in *Commutative Algebra* (Salvador, 1988), *Lect. Notes in Math.* No. 1430, (Berlin: Springer-Verlag) (1990)
- [K] Kunz E, Kähler Differentials, *Vieweg Adv. Stud. Math.* (Vieweg, Wiesbaden) (1986)
- [M] Matsumura H, *Commutative Algebra* (New York: Benjamin) (1980)
- [PS] Peskine C and Szpiro L, Liaison des variétés algébriques I, *Invent. Math.* **26** (1974) 271–302
- [Sh] Shafarevich I R, *Basic Algebraic Geometry*, (Berlin-Heidelberg-New York: Springer-Verlag) (1977)



# On the Ramanujan-Petersson conjecture for modular forms of half-integral weight

WINFRIED KOHNEN

Max-Planck-Institut für Mathematik, Gottfried-Claren-Str. 26, 53225 Bonn, Germany

MS received 10 May 1993

**Abstract.** It is shown that a “character twist” of a growth estimate for certain weighted infinite sums of Kloosterman sums which is equivalent to the Ramanujan-Petersson conjecture for modular forms of half-integral weight, can easily be proved using Deligne’s theorem (previously the Ramanujan-Petersson conjecture for modular forms of integral weight).

**Keywords.** Ramanujan-Petersson conjecture; modular forms of half-integral weight; Kloosterman sum.

## 1. Introduction

Let  $f$  be a cusp form of half-integral weight  $k + 1/2 \geq 3/2$  on the group  $\Gamma_0(4N)$  ( $N \in \mathbb{N}$ ) and denote by  $a(n)$  ( $n \in \mathbb{N}$ ) its Fourier coefficients. Then one conjectures that

$$a(n) \ll_{f,\varepsilon} n^{(k-1/2)/2+\varepsilon} \quad (\varepsilon > 0) \quad (1)$$

(if the weight is  $3/2$ , one also has to assume that  $f$  lies in the orthogonal complement of the space of theta functions [3, §4(C)]).

In analogy with Deligne’s theorem, previously the Ramanujan-Petersson conjecture, estimate (1) is known as the Ramanujan-Petersson conjecture for modular forms of half-integral weight.

Suppose in addition that  $f$  is a Hecke eigenform and denote by  $F$  a cuspidal Hecke eigenform of weight  $2k$  corresponding to  $f$  under the Shimura correspondence [3]. Then the ratio  $a(n):a(m)$  can be expressed in terms of the Hecke eigenvalues of  $F$  if  $n:m$  is a perfect square. It follows easily from this that—granting Deligne’s theorem for  $F$ —it is sufficient to prove (1) for  $n$  squarefree, and conversely that if we knew (1) to hold for all  $n$ , this would imply Deligne’s theorem for  $F$  (the latter fact had been observed independently by several authors).

Suppose that  $f$  and  $F$  are as above and  $n$  is squarefree. Then (1) by Waldspurger’s results [4] is equivalent to the bound

$$L_F(k, \chi_n) \ll_{F,\varepsilon} n^\varepsilon, \quad (\varepsilon > 0) \quad (2)$$

where  $\chi_n$  is the quadratic character of the field extension  $\varphi(\sqrt{(-1)^k n})/\varphi$  and  $L_F(s, \chi_n)$  denotes the Hecke  $L$ -function of  $F$  twisted with  $\chi_n$  (note that  $s = k$  is the point of symmetry of the functional equation of  $L_F(s, \chi_n)$ ). Inequality (2) can be viewed as a

generalization of the well-known conjectural bound

$$L\left(\frac{1}{2}, \chi_n\right) \ll_{\varepsilon} n^{\varepsilon} \quad (\varepsilon > 0)$$

( $L(s, \chi_n)$ —is the Dirichlet  $L$ -function attached to  $\chi_n$ ) and so far seems to be the main motivation why one should expect (1).

For simplicity let us suppose that  $k$  is even and  $k \geq 2$ . Then the purpose of this note is to point out some formulae which eventually could be interpreted to give some other kind of support for (1). Firstly, using the formalism of Poincaré series of half-integral weight and the explicit evaluation of their Fourier coefficients as in [2] (cf. also [1; §§ 2, 3]), we shall illustrate that (1) in case  $n$  is squarefree—for all forms  $f$  of a given weight and level—is equivalent to an inequality

$$s_{n,k} \ll_{k,\varepsilon} n^{\varepsilon} \quad (\varepsilon > 0; n \text{ squarefree})$$

where  $s_{n,k}$  is a certain infinite sum which is “twisted” with essentially a quadratic character attached to  $n$  and whose general term is the product of a Bessel function of order  $k - 1/2$  and a finite “elementary” exponential sum (both depending on  $n$ ). The latter more or less is well-known to experts, cf. e.g. [1]. Secondly, we will illustrate that the estimate

$$s_{n,k}^* \ll_{k,\varepsilon} n^{\varepsilon} \quad (\varepsilon > 0; n \text{ squarefree}). \quad (3)$$

where  $s_{n,k}^*$  defined in the same way as  $s_{n,k}$  but without “twisting”, in fact, is true. Using essentially well-known formulae in the literature, inequality (3) can be easily deduced from Deligne’s theorem.

For convenience, we will give all statements only in the simplest case where  $N = 1$ . Also instead of working with the full space of cusp forms of weight  $k + 1/2$  and level 4, we shall only consider the subspace  $S_{k+1/2}^+$  consisting of forms whose  $n$ th Fourier coefficients vanish unless  $n \equiv 0, 1 \pmod{4}$  [2].

## 2. Estimates for certain infinite sums

Throughout we shall suppose that  $k$  is even. We first recall some facts on Poincaré series of half-integral weight; for details the reader is referred to [2].

For  $m \in \mathbb{N}$  with  $m \equiv 0, 1 \pmod{4}$  let  $P_{k,m}$  be the  $m$ th Poincaré series in  $S_{k+1/2}^+$  characterized by

$$\langle f, P_{k,m} \rangle = C_k m^{-k+1/2} a(m) \quad (\forall f \in S_{k+1/2}^+), \quad (4)$$

where  $\langle, \rangle$  is the usual Petersson scalar product (normalized as in [27]),

$$C_k := \frac{1}{6} \cdot \Gamma(k - 1/2) (4\pi)^{-k+1/2}$$

and  $a(m)$  is the  $m$ th Fourier coefficient of  $f$ . Denote by  $g_{k,m}(n)$  ( $n \in \mathbb{N}$ ) the Fourier coefficients of  $P_{k,m}$ .

Let  $f \in S_{k+1/2}^+ \setminus \{0\}$  and let  $\{f_1, \dots, f_g\}$  be an orthonormal basis of  $S_{k+1/2}^+$  with  $f_1 = f/\|f\|$ . Let  $a_v(n)$  ( $n \in \mathbb{N}$ ) be the Fourier coefficients of  $f_v$ . Then from (4) we find

that

$$P_{k,m} = C_k m^{-k+1/2} \sum_{v=1}^g \overline{a_v(m)} f_v.$$

Taking the  $m$ th Fourier coefficients on both sides we obtain the Petersson formula

$$g_{k,m}(m) = C_k m^{-k+1/2} \sum_{v=1}^g |a_v(m)|^2. \quad (5)$$

Now let  $D$  be a positive fundamental discriminant and denote by

$$S_{k,D}: S_{k+1/2}^+ \rightarrow S_{2k}, \quad \sum_{n \geq 1} a(n) q^n \leftrightarrow \sum_{n \geq 1} \left( \sum_{d|n} \left( \frac{D}{d} \right) d^{k-1} a\left( \frac{n^2}{d^2} D \right) \right) q^n \quad (6)$$

$(q = e^{2\pi iz}; z \in H = \text{upper half-plane})$

the  $D$ th Shimura lift, where  $S_{2k}$  is the space of cusp forms of weight  $2k$  on  $SL_2(\mathbb{Z})$  [3, 2]. It was proved in [2; Theorems 1 and 2] that

$$S_{k,D}(P_{k,D}) = C'_k D^{k-1/2} F_{k,D} \quad (7)$$

where  $C'_k$  is a numerical constant not depending on  $D$  and

$$F_{k,D}(z) := \sum_{\substack{a,b,c \in \mathbb{Z} \\ b^2 - 4ac = D^2}} \omega_D(a, b, c) (az^2 + bz + c)^{-k} \quad (z \in H);$$

here  $\omega_D(a, b, c)$  is zero if  $(a, b, c, D) > 1$  and otherwise is  $(D/r)$  where  $r \in \mathbb{Z}$ ,  $(r, D) = 1$  and  $r$  is represented by the binary quadratic form with coefficients  $a, b$  and  $c$  (this definition does not depend on the choice of  $r$ ).

Moreover, it was shown in the course of the proof of Theorem 1 in [2] (loc. cit., Propos. 2) that the function  $F_{k,D}$  has the expansion

$$F_{k,D}(z) = \sum_{n \geq 1} c_{k,D}(n) q^n$$

with

$$c_{k,D}(n) = \frac{2(2\pi)^k}{(k-1)!} (n/D)^{k-1} \left[ (-1)^{k/2} D^{-1/2} \left( \frac{D}{n} \right) + \pi \sqrt{2} (n/D)^{1/2} \sum_{a \geq 1} a^{-1/2} \left( \sum_{\substack{b(2a) \\ b^2 \equiv D^2(4a)}} \omega_D\left(a, b, \frac{b^2 - D^2}{4a}\right) \exp\left(\frac{\pi i n b}{a}\right) \right) J_{k-1/2}\left(\frac{\pi n D}{a}\right) \right]; \quad (8)$$

( $J_{k-1/2}$  = Bessel function of order  $k - 1/2$ ).

Taking  $n = 1$  on the right-hand side of (6) and using (7) we find that

$$g_{k,D}(D) = C'_k D^{k-1/2} c_{k,D}(1),$$

hence by (8) we have

$$g_{k,D}(D) = C'_k \frac{2(2\pi)^k}{(k-1)!} (-1)^{k/2} + C'_k \frac{2(2\pi)^k}{(k-1)!} \pi \sqrt{2} s_{D,k} \quad (9)$$

where

$$s_{D,k} := \sum_{a \geq 1} a^{-1/2} \left( \sum_{\substack{b(2a) \\ b^2 \equiv D^2(4a)}} \omega_D \left( a, b, \frac{b^2 - D^2}{4a} \right) \exp \left( \frac{\pi i b}{a} \right) \right) J_{k-1/2} \left( \frac{\pi D}{a} \right).$$

Note that if

$$(a, D) = 1 \text{ then } \omega_D \left( a, b, \frac{b^2 - D^2}{4a} \right) = \left( \frac{D}{a} \right); \text{ also, if } (2a, D) = 1,$$

then in the  $a$ th term of  $s_{D,k}$  we can replace  $b$  by  $Db$ . Hence for  $(2a, D) = 1$  the  $a$ th term of  $s_{D,k}$  can more conveniently be written as

$$\left( \frac{D}{a} \right) a^{-1/2} \left( \sum_{\substack{b(2a) \\ b^2 \equiv 1(4a)}} \exp \left( \frac{\pi i Db}{a} \right) \right) J_{k-1/2} \left( \frac{\pi D}{a} \right).$$

From (5) and (9) we infer

### PROPOSITION 1

*The estimate*

$$a(D) \ll_{f,\varepsilon} D^{k/2 - 1/4 + \varepsilon} \quad (\varepsilon > 0; D \text{ a fundamental discriminant})$$

(Ramanujan–Petersson conjecture) holds for all  $f = \sum_{n \geq 1, n \equiv 0, 1(4)} a(n) q^n \in S_{k+1/2}^+$  if and only if the estimate

$$s_{D,k} \ll_{k,\varepsilon} D^\varepsilon \quad (\varepsilon > 0; D \text{ a fundamental discriminant})$$

is true.

A result similar to Proposition 1 can also be deduced from the arguments given in [1; §§ 2, 3].

For a positive fundamental discriminant  $D$  set

$$s_{D,K}^* := \sum_{a \geq 1} a^{-1/2} \left( \sum_{\substack{b(2a) \\ b^2 \equiv 1(4a)}} \exp \left( \frac{\pi i Db}{a} \right) \right) J_{k-1/2} \left( \frac{\pi D}{a} \right). \quad (10)$$

### PROPOSITION 2.

*One has*

$$s_{D,K}^* \ll_{k,\varepsilon} D^\varepsilon \quad (\varepsilon > 0). \quad (11)$$

*Proof.* The function  $F_{k,1}$  is a cusp form of weight  $2k$  on  $SL_2(\mathbb{Z})$ , hence Deligne's bound for its Fourier coefficients is valid. Therefore, in particular, we have

$$c_{k,1}(D) \ll_{k,\varepsilon} D^{k-1/2+\varepsilon} \quad (\varepsilon > 0).$$

Taking into account that  $\omega_1 = 1$  we find from (8) that

$$c_{k,1}(D) = \frac{2(2\pi)^k}{(k-1)!} D^{k-1} [(-1)^{k/2} + \pi \sqrt{2} D^{1/2} s_{D,k}^*].$$

This implies (11).

*Remarks.* i) If on the right-hand side of (10) one takes absolute values and uses the bound

$$\#\{b(\bmod 2a) | b^2 \equiv 1(\bmod 4a)\} \ll_{\varepsilon} a^{\varepsilon} \quad (\varepsilon > 0)$$

and well-known bounds for the Bessel functions, one obtains in a standard manner that

$$s_{D,k}^* \ll_{k,\varepsilon} D^{1/2+\varepsilon} \quad (\varepsilon > 0). \quad (12)$$

Note that (12) corresponds to the trivial bound  $\ll D^{k+\varepsilon}$  (even worse than the Hecke bound) for the Fourier coefficients  $c_{k,1}(D)$ . This shows that lots of cancellations must occur in the terms on the right-hand side of (10) in order to make (11) valid. Note that  $J_{k-1/2}$  is an elementary function and can be expressed in terms of polynomials and an exponential function.

ii) It seems to be very hard to get non-trivial bounds on  $s_{D,k}$  or  $s_{D,k}^*$  by trying to estimate the sums in a more or less direct way, e.g. by employing methods from elementary or analytic number theory or algebraic geometry (Weil's bound for Kloosterman sums). The best result in this direction is due to Iwaniec [1] who proved using very sophisticated arguments that

$$s_{D,k} \ll_{k,\varepsilon} D^{3/7+\varepsilon} \quad (\varepsilon > 0)$$

(as least if  $D \equiv 1 \pmod{4}$ ); the other case can also be deduced from the reasoning in [1]).

## References

- [1] Iwaniec H, Fourier coefficients of modular forms of half-integral weight. *Invent. Math.* **87** (1987) 385–401
- [2] Kohnen W, Fourier coefficients of modular forms of half-integral weight, *Math. Ann.* **271** (1985) 237–268
- [3] Shimura G, On modular forms of half-integral weight, *Ann. Math.* **97** (1973) 440–481
- [4] Waldspurger J-L, Sur les coefficients de Fourier des formes modulaires de poids demi-entier, *J. Math. Pures Appl.* **60** (1981) 375–484





# On composition of some general fractional integral operators

K C GUPTA and R C SONI

Department of Mathematics, M. R. Engineering College, Jaipur 302017, India

MS received 1 July 1991; revised 19 August 1993

**Abstract.** In the present paper we derive three interesting expressions for the composition of two most general fractional integral operators whose kernels involve the product of a general class of polynomials and a multivariable  $H$ -function. By suitably specializing the coefficients and the parameters in these functions we can get a large number of (new and known) interesting expressions for the composition of fractional integral operators involving classical orthogonal polynomials and simpler special functions (involving one or more variables) which occur rather frequently in problems of mathematical physics. We have mentioned here two special cases of the first composition formula. The first involves product of a general class of polynomials and the Fox's  $H$ -functions and is of interest in itself. The findings of Buschman [1] and Erdélyi [4] follow as simple special cases of this composition formula. The second special case involves product of the Jacobi polynomials, the Hermite polynomials and the product of two multivariable  $H$ -functions. The present study unifies and extends a large number of results lying scattered in the literature. Its findings are general and deep.

**Keywords.** Fractional integral operator; general class of polynomials; multivariable  $H$ -function.

## 1. Introduction

Fractional integral operators play an important role in the theory of integral equations and in problems of mathematical physics. We shall study in this paper the composition of fractional integral operators defined by means of the following equations

$$\begin{aligned} & R^{\eta, \alpha; m, n, v; N, P, Q, M', N', P', Q', \dots, M^{(r)}, N^{(r)}, P^{(r)}, Q^{(r)}, v_1, \dots, v_r} \\ & \quad x; e; z_1, \dots, z_r, a_j, \alpha'_j, \dots, \alpha_j^{(r)}, b_j, \beta'_j, \dots, \beta_j^{(r)}, c'_j, \gamma'_j, d'_j, \delta'_j, \dots, c_j^{(r)}, \gamma_j^{(r)}, d_j^{(r)}, \delta_j^{(r)} [f(x)] \\ &= x^{-\eta-\alpha-1} \int_0^x t^\eta (x-t)^\alpha S_n^m \left[ e \left( 1 - \frac{t}{x} \right)^v \right] H \left[ z_1 \left( 1 - \frac{t}{x} \right)^{v_1}, \dots, z_r \left( 1 - \frac{t}{x} \right)^{v_r} \right] f(t) dt. \end{aligned} \quad (1)$$

$$\begin{aligned} & W^{\eta, \alpha; m, n, v; N, P, Q, M', N', P', Q', \dots, M^{(r)}, N^{(r)}, P^{(r)}, Q^{(r)}, v_1, \dots, v_r} \\ & \quad x; e; z_1, \dots, z_r, a_j, \alpha'_j, \dots, \alpha_j^{(r)}, b_j, \beta'_j, \dots, \beta_j^{(r)}, c'_j, \gamma'_j, d'_j, \delta'_j, \dots, c_j^{(r)}, \gamma_j^{(r)}, d_j^{(r)}, \delta_j^{(r)} [f(x)] \\ &= x^\eta \int_x^\infty t^{-\eta-\alpha-1} (t-x)^\alpha S_n^m \left[ e \left( 1 - \frac{x}{t} \right)^v \right] H \left[ z_1 \left( 1 - \frac{x}{t} \right)^{v_1}, \dots, z_r \left( 1 - \frac{x}{t} \right)^{v_r} \right] f(t) dt \end{aligned} \quad (2)$$

Here  $S_n^m[x]$  denotes the general class of polynomials introduced by Srivastava

[15, p. 1, eqn. (1)]

$$S_n^m[x] = \sum_{k=0}^{[n/m]} \frac{(-n)_{mk}}{k!} A_{n,k} x^k, \quad n = 0, 1, 2, \dots, \quad (3)$$

where  $m$  is an arbitrary positive integer and the coefficients  $A_{n,k}$  ( $n, k \geq 0$ ) are arbitrary constants, real or complex. On suitably specializing the coefficients  $A_{n,k}$ ,  $S_n^m[x]$  yields a number of known polynomials as its special cases. These include, among others, the Hermite polynomials, the Jacobi polynomials, the Laguerre polynomials, the Bessel polynomials, the Gould-Hopper polynomials, the Brafman polynomials and several others [19, pp. 158–161].

The  $H$ -function of  $r$  complex variables  $z_1, \dots, z_r$  occurring in (1) and (2) was introduced by Srivastava and Panda [18]. We use here the following contracted form [17, p. 251, eqn. (C.1)] to denote it

$$H[z_1, \dots, z_r] = H_{\substack{0, N: M', N'; \dots; M^{(r)}, N^{(r)} \\ P, Q: P', Q'; \dots; P^{(r)}, Q^{(r)}}} \left[ \begin{matrix} z_1 \\ \vdots \\ z_r \end{matrix} \middle| \begin{matrix} (a_j; \alpha'_j, \dots, \alpha_j^{(r)})_{1, P}: (c'_j, \gamma'_j)_{1, P'}; \dots; (c_j^{(r)}, \gamma_j^{(r)})_{1, P^{(r)}} \\ (b_j; \beta'_j, \dots, \beta_j^{(r)})_{1, Q}: (d'_j, \delta'_j)_{1, Q'}; \dots; (d_j^{(r)}, \delta_j^{(r)})_{1, Q^{(r)}} \end{matrix} \right] \quad (4)$$

The defining integral and other details about this function can be found in the references given above.

It may be remarked here that all the Greek letters occurring in the right-hand side of (4) are assumed to be positive real numbers for standardization purposes; the definition of this function will, however, be meaningful even if some of these quantities are zero. Again, it is assumed throughout the present work that this function always satisfies the appropriate existence and convergence conditions of its defining integral [17, pp. 252–253, eqs (C.4–C.6)].

On account of the importance of the fractional integral operators in the theory of integral equations and other allied topics, these operators have been studied from time to time by a number of authors notably Riemann–Liouville [6], Weyl [6], Erdélyi [2, 3], Kober [11], Lowndes [12], Goyal and Jain [8], Gupta [9], Kalla [10], Saxena [13], Saxena and Kumbhat [14], Srivastava *et al* [16]. The importance of our study lies in the fact that the kernels (1) and (2) used here are most general in character.

To be specific, we shall assume throughout this paper that

$$f(x) = \begin{cases} O(|x|^\gamma), & |x| \rightarrow 0 \\ O(|x|^\delta e^{-\lambda|x|}), & |x| \rightarrow \infty \end{cases}$$

It is easy to verify that the operator defined by (1) exists if

- (i) The quantities  $v, v_1, \dots, v_r$  are all positive (some of them may decrease to zero provided that the resulting operator has a meaning).
- (ii)  $\operatorname{Re}(\alpha) + \sum_{i=1}^r v_i \min_{1 \leq j \leq M^{(i)}} [\operatorname{Re}(d_j^{(i)} / \delta_j^{(i)})] + 1 > 0$ .
- (iii)  $\operatorname{Re}(\eta + \gamma + 1) > 0$ .

and the operator defined by (2) exists if

$\operatorname{Re}(\lambda) > 0$  or  $\operatorname{Re}(\lambda) = 0$  and  $\operatorname{Re}(\eta - \delta) > 0$ , and the set of conditions (i) and (ii) specified for the existence of the operator (1) are satisfied.

## 2. Compositions of fractional integral operators

From the definition (1), we have  $R_x^{\eta, \alpha} \{R_x^{\theta, \beta} f(x)\}$

$$\begin{aligned}
 &= R \left\{ \begin{array}{l} \eta, \alpha; m, n, v; N, P, Q, M', N', P', Q', \dots, M^{(r)}, N^{(r)}, P^{(r)}, Q^{(r)}, v_1, \dots, v_r \\ x; e; z_1, \dots, z_r, a_j, \alpha_j', \dots, \alpha_j^{(r)}, b_j, \beta_j', \dots, \beta_j^{(r)}, c_j', \gamma_j', d_j', \delta_j', \dots, c_j^{(r)}, \gamma_j^{(r)}, d_j^{(r)}, \delta_j^{(r)} \\ \left\{ \begin{array}{l} R \quad \theta, \beta; m', n', v'; \quad N_1, P_1, Q_1, M^{(r+1)}, N^{(r+1)}, P^{(r+1)}, \\ x; e'; \quad z_{r+1}, \dots, z_{2r}, a_j', \alpha_j^{(r+1)}, \dots, \alpha_j^{(2r)}, b_j', \\ Q^{(r+1)}, \dots, M^{(2r)}, N^{(2r)}, P^{(2r)}, Q^{(2r)}, v_{r+1}, \dots, v_{2r} \\ \beta_j^{(r+1)}, \dots, \beta_j^{(2r)}, c_j^{(r+1)}, \gamma_j^{(r+1)}, d_j^{(r+1)}, \delta_j^{(r+1)}, \dots, c_j^{(2r)}, \gamma_j^{(2r)}, d_j^{(2r)}, \delta_j^{(2r)} \end{array} \right. f(x) \end{array} \right\} \\
 &= x^{-\eta-\alpha-1} \int_0^x (x-t)^\alpha S_n^m \left[ e \left( 1 - \frac{t}{x} \right)^v H \left[ z_1 \left( 1 - \frac{t}{x} \right)^{v_1}, \dots, z_r \left( 1 - \frac{t}{x} \right)^{v_r} \right] t^{\eta-\theta-\beta-1} \times \right. \\
 &\quad \left. \left\{ \int_0^t (t-\zeta)^\beta S_n^{m'} \left[ e' \left( 1 - \frac{\zeta}{t} \right)^{v'} \right] H \left[ z_{r+1} \left( 1 - \frac{\zeta}{t} \right)^{v_{r+1}}, \dots, z_{2r} \left( 1 - \frac{\zeta}{t} \right)^{v_{2r}} \right] \zeta^\theta f(\zeta) d\zeta \right\} dt \right] \\
 &\hspace{15em} (5)
 \end{aligned}$$

On changing the order of  $\zeta, t$ -integrals in the right-hand side of (5), we arrive at the following result after a little simplification

$$R_x^{\eta, \alpha} \{R_x^{\theta, \beta} f(x)\} = x^{-\eta-\alpha-1} \int_0^x \zeta^\theta f(\zeta) \Delta d\zeta \quad (6)$$

where

$$\begin{aligned}
 \Delta &= \int_\zeta^x (x-t)^\alpha (t-\zeta)^\beta t^{\eta-\theta-\beta-1} S_n^m \left[ e \left( 1 - \frac{t}{x} \right)^v \right] S_n^{m'} \left[ e' \left( 1 - \frac{\zeta}{t} \right)^{v'} \right] \\
 &\quad H \left[ z_1 \left( 1 - \frac{t}{x} \right)^{v_1}, \dots, z_r \left( 1 - \frac{t}{x} \right)^{v_r} \right] H \left[ z_{r+1} \left( 1 - \frac{\zeta}{t} \right)^{v_{r+1}}, \dots, z_{2r} \left( 1 - \frac{\zeta}{t} \right)^{v_{2r}} \right] dt \quad (7)
 \end{aligned}$$

The change in the order of integration in the right-hand side of (5) is justified by the well-known Fubini's theorem, provided that

$$\operatorname{Re}(\eta - \theta - \beta) > 0, \quad \operatorname{Re}(\theta + \gamma + 1) > 0, \quad \operatorname{Re}(\alpha) + \sum_{i=1}^r v_i \min_{1 \leq j \leq M^{(i)}} [\operatorname{Re}(d_j^{(i)} / \delta_j^{(i)})] + 1 > 0,$$

$$\operatorname{Re}(\beta) + \sum_{i=r+1}^{2r} v_i \min_{1 \leq j \leq M^{(i)}} [\operatorname{Re}(d_j^{(i)} / \delta_j^{(i)})] + 1 > 0.$$

Now to evaluate the value of  $\Delta$ , we express the general class of polynomials involved in the right-hand side of (7) in the series form and the multivariable  $H$ -functions in terms of their well known Mellin-Barnes contour integrals [17, pp. 251–252, eqns. (C.1)–(C.3)] and interchange the order of summation and integration in the result

thus obtain a result after a little simplification

$$\Delta = \sum_{k=0}^{[n/m]} \sum_{k'=0}^{[n'/m']} \frac{(-n)_{mk} (-n')_{m'k'}}{k! k'} A_{n,k} A'_{n',k'} e^k e'^{k'}$$

$$\frac{1}{(2\pi\omega)^{2r}} \int_{L_1} \dots \int_{L_{2r}} \phi_1(\xi_1) \dots \phi_{2r}(\xi_{2r}) \psi(\xi_1, \dots, \xi_r) \psi'(\xi_{r+1}, \dots, \xi_{2r}) z_1^{\xi_1} \dots z_{2r}^{\xi_{2r}} d\xi_1 \dots d\xi_{2r}$$

$$\int_{\zeta}^x (x-t)^{\alpha} (t-\zeta)^{\beta} t^{\eta-\theta-\beta-1} \left(1-\frac{t}{x}\right)^{vk+v_1\xi_1+\dots+v_r\xi_r} \left(1-\frac{\zeta}{t}\right)^{v'k'+v_{r+1}\xi_{r+1}+\dots+v_{2r}\xi_{2r}} dt \quad (8)$$

The  $t$ -integral occurring in the right-hand side of (8) is now evaluated by making the substitution  $t = x - (x - \zeta)w$  in it, further  ${}_2F_1$  thus obtained is expressed in terms of its well known Mellin-Barnes contour integral and the result thus obtained reinterpreted in terms of the  $H$ -function of  $2r+1$  variables to yield the value of  $\Delta$ . Now upon substituting the value of  $\Delta$  thus obtained in (6), we finally arrive at the following result after a little simplification

$$R_x^{\eta, \alpha} \left\{ R_x^{\theta, \beta} f(x) \right\} = x^{-\theta-\alpha-\beta-2} \sum_{k=0}^{[n/m]} \sum_{k'=0}^{[n'/m']} \frac{(-n)_{mk} (-n')_{m'k'}}{k! k'} A_{n,k} A'_{n',k'} e^k e'^{k'}$$

$$x^{-(vk+v'k')} \int_0^x \zeta^{\theta} (x-\zeta)^{\alpha+\beta+vk+v'k'+1} H_{P+P_1+3, Q+Q_1+2; P', Q'; \dots; P^{(2r)}, Q^{(2r)}; 0, 1}$$

$$\left[ \begin{array}{c} z_1 \left(1 - \frac{\zeta}{x}\right)^{v_1} \\ \vdots \\ z_{2r} \left(1 - \frac{\zeta}{x}\right)^{v_{2r}} \\ - \left(1 - \frac{\zeta}{x}\right) \end{array} \right] \left( \begin{array}{c} \eta - \theta - \beta - v'k'; \frac{0, \dots, 0}{r}, v_{r+1}, \dots, v_{2r}, 1 \\ -1 - \alpha - \beta - vk - v'k'; v_1, \dots, v_{2r}, 1 \end{array} \right),$$

$$\left( -\alpha - vk; v_1, \dots, v_r, \frac{0, \dots, 0}{r}, 1 \right), \left( -\beta - v'k'; \frac{0, \dots, 0}{r}, v_{r+1}, \dots, v_{2r}, 0 \right),$$

$$\left( \eta - \theta - \beta - v'k'; \frac{0, \dots, 0}{r}, v_{r+1}, \dots, v_{2r}, 0 \right),$$

$$\left( a_j; \alpha'_j, \dots, \alpha_j^{(r)}, \frac{0, \dots, 0}{r+1} \right)_{1,N}, \left( \alpha'_j; \frac{0, \dots, 0}{r}, \alpha_j^{(r+1)}, \dots, \alpha_j^{(2r)}, 0 \right)_{1,P_1}, \left( a_j; \alpha'_j, \dots, \alpha_j^{(r)}, \frac{0, \dots, 0}{r+1} \right)_{N+1,P} :$$

$$\left( b_j; \beta'_j, \dots, \beta_j^{(r)}, \frac{0, \dots, 0}{r+1} \right)_{1,Q}, \left( \beta'_j; \frac{0, \dots, 0}{r}, \beta_j^{(r+1)}, \dots, \beta_j^{(2r)}, 0 \right)_{1,Q_1} :$$

$$\left[ (c'_j, \gamma'_j)_{1,P'}; \dots; (c_j^{(2r)}, \gamma_j^{(2r)})_{1,P(2r)}; - \right. \\ \left. (d'_j, \delta'_j)_{1,Q'}; \dots; (d_j^{(2r)}, \delta_j^{(2r)})_{1,Q(2r)}; (0, 1) \right] f(\zeta) d\zeta \quad (9)$$

$\frac{1}{r}$  would mean  $r$  zeros, and so on, and the conditions necessary for the change of order of integration stated earlier are satisfied.

On following the procedure adopted by Erdélyi [4] and on making the use of well-known transformation formula for the Gauss' hypergeometric function [5, p. 64, eqn. (23)], the following commutative property of the operator given by (1) can be easily established

$$R_x^{\eta, \alpha} \{ R_x^{\theta, \beta} f(x) \} = R_x^{\theta, \beta} \{ R_x^{\eta, \alpha} f(x) \}.$$

The expression for the composition of the fractional integral operator defined by (2), can be easily derived in a similar manner. We have

$$\begin{aligned} & W^{\eta, \alpha; m, n, v; N, P, Q, M', N', P', Q', \dots, M^{(r)}, N^{(r)}, P^{(r)}, Q^{(r)}, v_1, \dots, v_r} \\ & \quad x; e; z_1, \dots, z_r, a_j, \alpha_j', \dots, \alpha_j^{(r)}, b_j, \beta_j', \dots, \beta_j^{(r)}, c_j, \gamma_j', d_j', \dots, c_j^{(r)}, \gamma_j^{(r)}, d_j^{(r)}, \delta_j^{(r)} \\ & \left\{ W^{\theta, \beta; m', n', v'; N_1, P_1, Q_1, M^{(r+1)}, N^{(r+1)}, P^{(r+1)}, Q^{(r+1)}, \right. \\ & \quad x; e'; \quad z_{r+1}, \dots, z_{2r}, a_j', \alpha_j^{(r+1)}, \dots, \alpha_j^{(2r)}, b_j', \beta_j^{(r+1)}, \\ & \quad \dots, M^{(2r)}, N^{(2r)}, P^{(2r)}, Q^{(2r)}, v_{r+1}, \dots, v_{2r} \\ & \quad \dots, \beta_j^{(2r)}, c_j^{(r+1)}, \gamma_j^{(r+1)}, d_j^{(r+1)}, \delta_j^{(r+1)}, \dots, c_j^{(2r)}, \gamma_j^{(2r)}, d_j^{(2r)}, \delta_j^{(2r)} f(x) \left. \right\} \\ & = W^{\theta, \beta; m', n', v'; N_1, P_1, Q_1, M^{(r+1)}, N^{(r+1)}, P^{(r+1)}, Q^{(r+1)}, \\ & \quad x; e'; z_{r+1}, \dots, z_{2r}, a_j', \alpha_j^{(r+1)}, \dots, \alpha_j^{(2r)}, b_j', \beta_j^{(r+1)}, \\ & \quad \dots, M^{(2r)}, N^{(2r)}, P^{(2r)}, Q^{(2r)}, v_{r+1}, \dots, v_{2r} \\ & \quad \dots, \beta_j^{(2r)}, c_j^{(r+1)}, \gamma_j^{(r+1)}, d_j^{(r+1)}, \delta_j^{(r+1)}, \dots, c_j^{(2r)}, \gamma_j^{(2r)}, d_j^{(2r)}, \delta_j^{(2r)} \\ & \left\{ W^{\eta, \alpha; m, n, v; N, P, Q, M', N', P', Q', \dots, M^{(r)}, N^{(r)}, P^{(r)}, Q^{(r)}, v_1, \dots, v_r} \right. \\ & \quad x; e; z_1, \dots, z_r, a_j, \alpha_j', \dots, \alpha_j^{(r)}, b_j, \beta_j', \dots, \beta_j^{(r)}, c_j, \gamma_j', d_j', \dots, c_j^{(r)}, \gamma_j^{(r)}, d_j^{(r)}, \delta_j^{(r)} f(x) \left. \right\} \\ & = x^\eta \sum_{k=0}^{[n/m]} \sum_{k'=0}^{[n'/m']} \frac{(-n)_{mk} (-n')_{m'k'}}{k! k'} A_{n,k} A_{n',k'} e^k e^{k'} \\ & \int_x^\infty \zeta^{-\eta-\alpha-\beta-vk-v'k'-2} (\zeta-x)^{\alpha+\beta+vk+v'k'+1} \\ & \quad H^{0, N+N_1+3; M', N'; \dots; M^{(2r)}, N^{(2r)}; 1, 0} \\ & \quad P+P_1+3, Q+Q_1+2; P', Q'; \dots; P^{(2r)}, Q^{(2r)}; 0, 1 \\ & \quad \left[ \begin{array}{c} z_1 \left( 1 - \frac{x}{\zeta} \right)^{v_1} \\ \vdots \\ z_{2r} \left( 1 - \frac{x}{\zeta} \right)^{v_{2r}} \\ - \left( 1 - \frac{x}{\zeta} \right) \end{array} \right] \\ & \left( \theta - \eta - \alpha - vk; v_1, \dots, v_r, \frac{0, \dots, 0}{r}, 1 \right), \left( -\beta - v'k'; \frac{0, \dots, 0}{r}, v_{r+1}, \dots, v_{2r}, 1 \right), \end{aligned}$$

$$\begin{aligned}
& \left( \theta - \eta - \alpha - \nu k; v_1, \dots, v_r, \frac{0, \dots, 0}{r+1} \right), \\
& \left( a_j; \alpha'_j, \dots, \alpha_j^{(r)}, \frac{0, \dots, 0}{r+1} \right)_{1,N}, \left( a'_j; \frac{0, \dots, 0}{r}, \alpha_j^{(r+1)}, \dots, \alpha_j^{(2r)}, 0 \right)_{1,P_1}, \left( a_j; \alpha'_j, \dots, \alpha_j^{(r)}, \frac{0, \dots, 0}{r+1} \right)_{N+1,P} : \\
& \left( b_j; \beta'_j, \dots, \beta_j^{(r)}, \frac{0, \dots, 0}{r+1} \right)_{1,Q}, \left( b'_j; \frac{0, \dots, 0}{r}, \beta_j^{(r+1)}, \dots, \beta_j^{(2r)}, 0 \right)_{1,Q_1} : \\
& \left[ \begin{aligned} & (c'_j, \gamma'_j)_{1,P'}; \dots; (c_j^{(2r)}, \gamma_j^{(2r)})_{1,P^{(2r)}}; \dots \\ & (d'_j, \delta'_j)_{1,Q'}; \dots; (d_j^{(2r)}, \delta_j^{(2r)})_{1,Q^{(2r)}}; (0, 1) \end{aligned} \right] f(\zeta) d\zeta \quad (10)
\end{aligned}$$

provided that the composite operator defined by the left-hand side of (10) exists and the following conditions are satisfied

$$\operatorname{Re}(\lambda) > 0 \text{ or } \operatorname{Re}(\lambda) = 0 \text{ and } \min[\operatorname{Re}(\eta - \theta, \theta - \delta, \theta + \beta - \delta)] > 0,$$

$$\operatorname{Re}(\alpha) + \sum_{i=1}^r v_i \min_{1 \leq j \leq M^{(i)}} [\operatorname{Re}(d_j^{(i)} / \delta_j^{(i)})] + 1 > 0,$$

$$\operatorname{Re}(\beta) + \sum_{i=r+1}^{2r} v_i \min_{1 \leq j \leq M^{(i)}} [\operatorname{Re}(d_j^{(i)} / \delta_j^{(i)})] + 1 > 0.$$

Finally, we obtain the expression for the composition of the fractional integral operators defined by (1) and (2), we have  $R_x^{\eta, \alpha} \{ W_x^{\theta, \beta} f(x) \}$

$$\begin{aligned}
& = R^{\eta, \alpha; m, n, \nu; N, P, Q, M', N', P', Q', \dots, M^{(r)}, N^{(r)}, P^{(r)}, Q^{(r)}, v_1, \dots, v_r} \\
& \quad x; z_1, \dots, z_r, a_j, \alpha'_j, \dots, \alpha_j^{(r)}, b_j, \beta'_j, \dots, \beta_j^{(r)}, c'_j, \gamma'_j, d'_j, \delta'_j, \dots, c_j^{(r)}, \gamma_j^{(r)}, d_j^{(r)}, \delta_j^{(r)} \\
& \quad \left\{ W^{\theta, \beta; m', n', \nu';} \right. \\
& \quad \left. \begin{aligned} & x; e' \\ & N_1, P_1, Q_1, M^{(r+1)}, N^{(r+1)}, P^{(r+1)}, Q^{(r+1)}, \\ & z_{r+1}, \dots, z_{2r}, a'_j, \alpha_j^{(r+1)}, \dots, \alpha_j^{(2r)}, b'_j, \beta_j^{(r+1)}, \\ & \dots, M^{(2r)}, N^{(2r)}, P^{(2r)}, Q^{(2r)}, v_{r+1}, \dots, v_{2r} \\ & \dots, \beta_j^{(2r)}, c_j^{(r+1)}, \gamma_j^{(r+1)}, d_j^{(r+1)}, \delta_j^{(r+1)}, \dots, c_j^{(2r)}, \gamma_j^{(2r)}, d_j^{(2r)}, \delta_j^{(2r)} f(x) \end{aligned} \right\} \\
& = x^{-\eta - \alpha - 1} \int_0^x (x-t)^{\alpha} S_n^m \left[ e \left( 1 - \frac{t}{x} \right)^{\nu} \right] H \left[ z_1 \left( 1 - \frac{t}{x} \right)^{v_1}, \dots, z_r \left( 1 - \frac{t}{x} \right)^{v_r} \right] t^{\eta + \theta} \\
& \quad \left\{ \int_t^{\infty} (\zeta - t)^{\beta} S_n^m \left[ e' \left( 1 - \frac{t}{\zeta} \right)^{\nu'} \right] H \left[ z_{r+1} \left( 1 - \frac{t}{\zeta} \right)^{v_{r+1}}, \dots, z_{2r} \left( 1 - \frac{t}{\zeta} \right)^{v_{2r}} \right] \zeta^{-\theta - \beta - 1} f(\zeta) d\zeta \right\} dt \quad (11)
\end{aligned}$$

Now changing the order of  $\zeta, t$ -integrals in the right-hand side of (11), we get after a little simplification.

$$R_x^{\eta,\alpha} \{W_x^{\theta,\beta} f(x)\} = x^{-\eta-\alpha-1} \left[ \int_0^x \zeta^{-\theta-\beta-1} f(\zeta) \Delta_1 d\zeta + \int_x^\infty \zeta^{-\theta-\beta-1} f(\zeta) \Delta_2 d\zeta \right] \quad (12)$$

where

$$\begin{aligned} \Delta_1 = & \int_0^\zeta (x-t)^\alpha (\zeta-t)^\beta t^{\eta+\theta} S_n^m \left[ e \left( 1 - \frac{t}{x} \right)^v \right] S_{n'}^{m'} \left[ e' \left( 1 - \frac{t}{\zeta} \right)^{v'} \right] \\ & H \left[ z_1 \left( 1 - \frac{t}{x} \right)^{v_1}, \dots, z_r \left( 1 - \frac{t}{x} \right)^{v_r} \right] H \left[ z_{r+1} \left( 1 - \frac{t}{\zeta} \right)^{v_{r+1}}, \dots, z_{2r} \left( 1 - \frac{t}{\zeta} \right)^{v_{2r}} \right] dt \end{aligned} \quad (13)$$

and

$$\begin{aligned} \Delta_2 = & \int_0^x (x-t)^\alpha (\zeta-t)^\beta t^{\eta+\theta} S_n^m \left[ e \left( 1 - \frac{t}{x} \right)^v \right] S_{n'}^{m'} \left[ e' \left( 1 - \frac{t}{\zeta} \right)^{v'} \right] \\ & H \left[ z_1 \left( 1 - \frac{t}{x} \right)^{v_1}, \dots, z_r \left( 1 - \frac{t}{x} \right)^{v_r} \right] H \left[ z_{r+1} \left( 1 - \frac{t}{\zeta} \right)^{v_{r+1}}, \dots, z_{2r} \left( 1 - \frac{t}{\zeta} \right)^{v_{2r}} \right] dt \end{aligned} \quad (14)$$

The change in the order of integration in the right-hand side of (11) is justified by the well-known Fubini's theorem, provided that

$\operatorname{Re}(\eta + \theta + 1) > 0$ ,  $\operatorname{Re}(\gamma - \theta - \beta) > 0$ ,  $\operatorname{Re}(\lambda) > 0$  or  $\operatorname{Re}(\lambda) = 0$  and

$$\min[\operatorname{Re}(\theta - \delta, \theta + \beta - \delta)] > 0, \operatorname{Re}(\alpha) + \sum_{i=1}^r v_i \min_{1 \leq j \leq M^{(i)}} [\operatorname{Re}(d_j^{(i)} / \delta_j^{(i)})] + 1 > 0,$$

$$\operatorname{Re}(\beta) + \sum_{i=r+1}^{2r} v_i \min_{1 \leq j \leq M^{(i)}} [\operatorname{Re}(d_j^{(i)} / \delta_j^{(i)})] + 1 > 0.$$

Now evaluating the values of  $\Delta_1$  and  $\Delta_2$  by the usual procedure, we finally obtain the following result after a little simplification

$$R_x^{\eta,\alpha} \{W_x^{\theta,\beta} f(x)\} = \Gamma(\eta + \theta + 1) \sum_{k=0}^{[n/m]} \sum_{k'=0}^{[n'/m']} \frac{(-n)_{mk} (-n')_{m'k'}}{k! k'!} A_{n,k} A'_{n',k'} e^k e'^{k'}$$

$$\left\{ x^{-\eta-1} \int_0^x \zeta^\eta \left( 1 - \frac{\zeta}{x} \right)^{\alpha+\beta+vk+v'k'+1} \right.$$

$$\begin{aligned} & {}^H 0, N + N_1 + 2 : M', N'; \dots; M^{(2r)}, N^{(2r)}; 1, 0 \\ & P + P_1 + 2, Q + Q_1 + 2 : P', Q'; \dots; P^{(2r)}, Q^{(2r)}; 0, 1 \end{aligned} \left[ \begin{array}{c} z_1 \left( 1 - \frac{\zeta}{x} \right)^{v_1} \\ \vdots \\ z_{2r} \left( 1 - \frac{\zeta}{x} \right)^{v_{2r}} \\ - \left( \frac{\zeta}{x} \right) \end{array} \right]$$

$$\begin{aligned}
& \left( a_j; \alpha'_j, \dots, \alpha_j^{(r)}, \frac{0, \dots, 0}{r+1} \right)_{1,N}, \left( a'_j; \frac{0, \dots, 0}{r}, \alpha_j^{(r+1)}, \dots, \alpha_j^{(2r)}, 0 \right)_{1,P_1}, \left( a_j; \alpha'_j, \dots, \alpha_j^{(r)}, \frac{0, \dots, 0}{r+1} \right)_{N+1,P} : \\
& \left( -1 - \eta - \theta - \beta - v'k'; \frac{0, \dots, 0}{r}, v_{r+1}, \dots, v_{2r}, 1 \right), (-1 - \eta - \theta - \alpha - \beta - vk - v'k'; \\
& v_1, \dots, v_{2r}, 0), \left( b_j; \beta'_j, \dots, \beta_j^{(r)}, \frac{0, \dots, 0}{r+1} \right)_{1,Q}, \left( b'_j; \frac{0, \dots, 0}{r}, \beta_j^{(r+1)}, \dots, \beta_j^{(2r)}, 0 \right)_{1,Q_1} : \\
& \left[ \begin{aligned} & (c'_j, \gamma'_j)_{1,P}; \dots; (c_j^{(2r)}, \gamma_j^{(2r)})_{1,P(2r)}; - \\ & (d'_j, \delta'_j)_{1,Q}; \dots; (d_j^{(2r)}, \delta_j^{(2r)})_{1,Q(2r)}; (0, 1) \end{aligned} \right] f(\zeta) d\zeta \\
& + x^\theta \int_x^z \zeta^{-\theta-1} \left( 1 - \frac{x}{\zeta} \right)^{\alpha + \beta + vk + v'k' + 1} \\
& {}^H \begin{matrix} 0, N + N_1 + 2; M', N'; \dots; M^{(2r)}, N^{(2r)}; 1, 0 \\ P + P_1 + 2, Q + Q_1 + 2; P', Q'; \dots; P^{(2r)}, Q^{(2r)}; 0, 1 \end{matrix} \left[ \begin{matrix} z_1 \left( 1 - \frac{x}{\zeta} \right)^{v_1} \\ \vdots \\ z_{2r} \left( 1 - \frac{x}{\zeta} \right)^{v_{2r}} \\ - \left( \frac{x}{\zeta} \right) \end{matrix} \right] \\
& (-1 - \eta - \theta - \alpha - \beta - vk - v'k'; v_1, \dots, v_{2r}, 1), \left( -\alpha - vk; v_1, \dots, v_r, \frac{0, \dots, 0}{r}, 1 \right), \\
& \left( a_j; \alpha'_j, \dots, \alpha_j^{(r)}, \frac{0, \dots, 0}{r+1} \right)_{1,N}, \left( a'_j; \frac{0, \dots, 0}{r}, \alpha_j^{(r+1)}, \dots, \alpha_j^{(2r)}, 0 \right)_{1,P_1}, \left( a_j; \alpha'_j, \dots, \alpha_j^{(r)}, \frac{0, \dots, 0}{r+1} \right)_{N+1,P} : \\
& \left( -1 - \eta - \theta - \alpha - vk; v_1, \dots, v_r, \frac{0, \dots, 0}{r}, 1 \right), (-1 - \eta - \theta - \alpha - \beta - vk' - v'k'; v_1, \dots, v_{2r}, 0), \\
& \left( b_j; \beta'_j, \dots, \beta_j^{(r)}, \frac{0, \dots, 0}{r+1} \right)_{1,Q}, \left( b'_j; \frac{0, \dots, 0}{r}, \beta_j^{(r+1)}, \dots, \beta_j^{(2r)}, 0 \right)_{1,Q_1} : \\
& \left. \begin{aligned} & (c'_j, \gamma'_j)_{1,P}; \dots; (c_j^{(2r)}, \gamma_j^{(2r)})_{1,P(2r)}; - \\ & (d'_j, \delta'_j)_{1,Q}; \dots; (d_j^{(2r)}, \delta_j^{(2r)})_{1,Q(2r)}; (0, 1) \end{aligned} \right] f(\zeta) d\zeta \} \quad (15)
\end{aligned}$$

where the conditions of the existence of the operators and the conditions necessary for the change of order of integration stated earlier are satisfied.

The other composition of a mixed type can be handled similarly, indeed we have

$$R_x^{\eta, \alpha} \{ W_x^{\theta, \beta} f(x) \} = W_x^{\theta, \beta} \{ R_x^{\eta, \alpha} f(x) \}.$$



get the following result after a little simplification

$$\begin{aligned}
 & R \begin{matrix} \eta, \alpha; m, n, v; M, N, P, Q, v \\ x; e; z, c_j, \gamma_j, d_j, \delta_j \end{matrix} \left\{ R \begin{matrix} \theta, \beta; m', n', v'; M', N', P', Q', v' \\ x; e'; z', c'_j, \gamma'_j, d'_j, \delta'_j \end{matrix} f(x) \right\} \\
 &= x^{-\theta-\alpha-\beta-2} \sum_{k=0}^{[n/m]} \sum_{k'=0}^{[n'/m']} \frac{(-n)_{mk} (-n')_{m'k'}}{k! k'} A_{n,k} A'_{n',k} e^k e'^{k'} \\
 & \cdot x^{-(vk+v'k')} \int_0^x \zeta^\theta (x-\zeta)^{\alpha+\beta+vk+v'k'+1} \\
 & \quad H \begin{matrix} 0, 2; M, N; M', N'+1; 1, 0 \\ 2, 1; P, Q; P'+1, Q'+1; 0, 1 \end{matrix} \left[ \begin{matrix} z \left(1 - \frac{\zeta}{x}\right)^v \\ z' \left(1 - \frac{\zeta}{x}\right)^{v'} \\ - \left(1 - \frac{\zeta}{x}\right) \end{matrix} \right] \\
 & (\eta - \theta - \beta - v'k'; 0, v', 1), (-\alpha - vk; v, 0, 1): \\
 & (-1 - \alpha - \beta - vk - v'k'; v, v', 1) : \\
 & \left. \begin{matrix} (c_j, \gamma_j)_{1,p}; (-\beta - v'k', v'), (c'_j, \gamma'_j)_{1,p'}; - \\ (d_j, \delta_j)_{1,q}; (d'_j, \delta'_j)_{1,q'}, (\eta - \theta - \beta - v'k', v'); (0, 1) \end{matrix} \right] f(\zeta) d\zeta \quad (16)
 \end{aligned}$$

where

$$\begin{aligned}
 & R \begin{matrix} \eta, \alpha; m, n, v; M, N, P, Q, v \\ x; e; z, c_j, \gamma_j, d_j, \delta_j \end{matrix} [f(x)] = x^{-\eta-\alpha-1} \int_0^x t^\eta (x-t)^\alpha S_n^m \left[ e \left(1 - \frac{t}{x}\right)^v \right] \\
 & \quad H \begin{matrix} M, N \\ P, Q \end{matrix} \left[ z \left(1 - \frac{t}{x}\right)^v \left| \begin{matrix} (c_j, \gamma_j)_{1,p} \\ (d_j, \delta_j)_{1,q} \end{matrix} \right. \right] f(t) dt
 \end{aligned}$$

and the conditions of validity of (16) easily obtainable from those of (9) are satisfied.

The composition formula given by (16) is sufficiently general in nature and is of interest by itself. Thus by reducing the general class of polynomials and the Fox's  $H$ -functions occurring in it into their respective special cases [19, pp. 158–161; 17, pp. 18–19], we can obtain from it expressions for the composition of several fractional integral operators involving a large number of classical polynomials and simpler special functions. For example, if in the composition formula (16), we take  $n = n' = 0$  (the polynomials  $S_0^m$  and  $S_0^{m'}$  will reduce to  $A_{0,0}$  and  $A'_{0,0}$  respectively and can be taken to be unity without loss of generality) and further put  $M = Q = M' = Q' = 1$ ,  $N = P = N' = P' = 0$ ,  $d_j = d'_j = 0$ ,  $\delta_j = \delta'_j = 1$ ,  $v = v' = 0$  and let  $z \rightarrow 0$ ,  $z' \rightarrow 0$ , we get the results which are in essence same as obtained by Buschman [1, p. 100, eq. (2.4)] and Erdélyi [4, p. 166, eqn. (6.1)]. Also, the composition formulae (10) and (15) reduce to the remaining formulae given by Erdélyi [4, pp. 166–167, eqs (6.2) and (6.3)], if we specialize our operators to the simple operators studied by him.

(ii) If in the composition formula given by (9) we reduce the general class of polynomials  $S_n^m$  and  $S_{n'}^{m'}$  to the Jacobi polynomials and the Hermite polynomials respectively [19, p. 159, eqn. (1.6); p. 158, eqn. (1.4)], we arrive at the following new

$$\begin{aligned}
& x^{-\eta-x-1} \int_0^x (x-t)^\alpha P_n^{(\mu, \rho)} \left[ 1 - 2 \left( 1 - \frac{t}{x} \right) \right] H \left[ z_1 \left( 1 - \frac{t}{x} \right)^{v_1}, \dots, z_r \left( 1 - \frac{t}{x} \right)^{v_r} \right] t^{\eta-\theta-\beta-1} \\
& \left\{ \int_0^t (t-\zeta)^\beta \left( 1 - \frac{\zeta}{t} \right)^{n'/2} H_{n'} \left[ \frac{1}{2 \sqrt{\left( 1 - \frac{\zeta}{t} \right)}} \right] H \left[ z_{r+1} \left( 1 - \frac{\zeta}{t} \right)^{v_{r+1}}, \dots, \right. \right. \\
& \quad \left. \left. z_{2r} \left( 1 - \frac{\zeta}{t} \right)^{v_{2r}} \right] \zeta^\theta f(\zeta) d\zeta \right\} dt \\
& = x^{-\theta-x-\beta-2} \sum_{k=0}^n \sum_{k'=0}^{[n'/2]} \frac{(-n)_k (-n')_{2k'}}{k! k'} \binom{\eta+\mu}{n} \frac{(\mu+\rho+n+1)_k}{(\mu+1)_k} (-1)^{k'} \\
& x^{-(k+k')} \int_0^x \zeta^\theta (x-\zeta)^{\alpha+\beta+k+k'+1}
\end{aligned}$$

$$\begin{aligned}
& {}^H_{P+P_1+3, Q+Q_1+2; P', Q'; \dots; P^{(2r)}, Q^{(2r)}; 0, 1} \left[ \begin{matrix} z_1 \left( 1 - \frac{\zeta}{x} \right)^{v_1} \\ \vdots \\ z_{2r} \left( 1 - \frac{\zeta}{x} \right)^{v_{2r}} \\ - \left( 1 - \frac{\zeta}{x} \right) \end{matrix} \right] \\
& \left( \eta - \theta - \beta - k'; \frac{0, \dots, 0}{r}, v_{r+1}, \dots, v_{2r}, 1 \right), \left( -\alpha - k; v_1, \dots, v_r, \frac{0, \dots, 0}{r}, 1 \right), \\
& \left( -1 - \alpha - \beta - k - k'; v_1, \dots, v_{2r}, 1 \right), \\
& \left( -\beta - k'; \frac{0, \dots, 0}{r}, v_{r+1}, \dots, v_{2r}, 0 \right), \\
& \left( \eta - \theta - \beta - k'; \frac{0, \dots, 0}{r}, v_{r+1}, \dots, v_{2r}, 0 \right), \\
& \left( a_j; \alpha'_j, \dots, \alpha_j^{(r)}, \frac{0, \dots, 0}{r+1} \right)_{1, N}, \left( a'_j; \frac{0, \dots, 0}{r}, \alpha_j^{(r+1)}, \dots, \alpha_j^{(2r)}, 0 \right)_{1, P_1}, \left( a_j; \alpha'_j, \dots, \alpha_j^{(r)}, \frac{0, \dots, 0}{r+1} \right)_{N+1, P} : \\
& \left( b_j; \beta'_j, \dots, \beta_j^{(r)}, \frac{0, \dots, 0}{r+1} \right)_{1, Q}, \left( b'_j; \frac{0, \dots, 0}{r}, \beta_j^{(r+1)}, \dots, \beta_j^{(2r)}, 0 \right)_{1, Q_1} : \\
& \left. \left( c'_j, \gamma'_j \right)_{1, P'; \dots; (c_j^{(2r)}, \gamma_j^{(2r)})_{1, P^{(2r)}; -} \right] f(\zeta) d\zeta \quad (17) \\
& \left( d'_j, \delta'_j \right)_{1, Q'; \dots; (d_j^{(2r)}, \delta_j^{(2r)})_{1, Q^{(2r)}; (0, 1)}
\end{aligned}$$

provided that the various integrals involved in the left-hand side of (17) are absolutely convergent.

Several other interesting special cases of (9), (10) and (15) involving a large variety of polynomials (which are special cases of  $S_n^m$  and  $S_n^{m'}$ ) and numerous simpler special functions (which are particular cases of the multivariable  $H$ -function) can also be worked out but we do not record them here for lack of space.

## Acknowledgement

The authors are thankful to the referee for making useful suggestions.

## References

- [1] Buschman R G, Fractional integration, *Math. Jpn.* **9** (1964) 99–106
- [2] Erdélyi A, On fractional integration and its application to the theory of Hankel transforms, *Quart. J. Math. Oxford Ser. (2)* **11** (1940) 293–303
- [3] Erdélyi A, On some functional transformations, *Univ. Politec. Torino Rend. Sem. Mat.* **10** (1950–51) 217–234
- [4] Erdélyi A, Fractional integrals of generalized functions, in 'Fractional calculus and its applications' (*Lecture Notes in Math.*, Vol. 457), 151–170 New York, Springer-Verlag (1975)
- [5] Erdélyi A, Magnus W, Oberhettinger F and Tricomi F G, *Higher transcendental functions*, Vol. I, New York McGraw-Hill, (1953)
- [6] Erdélyi A, Magnus W, Oberhettinger F and Tricomi F G, *Tables of integral Transforms*, Vol. II New York, McGraw-Hill (1954)
- [7] Fox C, The G and H functions as symmetrical Fourier kernels, *Trans. Am. Math. Soc.* **98** (1961) 395–429
- [8] Goyal S P and Jain R M, Fractional integral operators and the generalized hypergeometric functions, *Indian J. Pure Appl. Math.*, **18** (1987) 251–259
- [9] Gupta R, Fractional integral operators and a general class of polynomials, *Indian J. Math.* **32** (1990) 69–77
- [10] Kalla S L, Operators of fractional integration, Analytic Functions, Kozubunik 1979, (Lecture Notes in Math., Vol. 798), 258–280, Berlin, Springer) (1980)
- [11] Kober H, On fractional integrals and derivatives, *Quart. J. Math. Oxford Ser. (2)* **11** (1940) 193–211
- [12] Lowndes J S, A generalization of the Erdélyi-Kober operators, *Proc. Edinburgh Math. Soc. Ser. (2)* **17** (1970) 139–148
- [13] Saxena R K, On fractional integration operators, *Math. Z.* **96** (1967) 288–291
- [14] Saxena R K and Kumbhat R K, Integral operators involving  $H$ -function, *Indian J. Pure Appl. Math.*, **5** (1974) 1–6
- [15] Srivastava H M, A contour integral involving Fox's  $H$ -function, *Indian J. Math.* **14** (1972) 1–6
- [16] Srivastava H M, Goyal S P and Jain R M, Fractional integral operators involving a general class of polynomials, *J. Math. Anal. Appl.* **148** (1990) 87–100
- [17] Srivastava H M, Gupta K C and Goyal S P, The  $H$ -functions of one and two variables with applications, South Asian Publishers, New Delhi/Madras, 1982
- [18] Srivastava H M and Panda R, Some bilateral generating functions for a class of generalized hypergeometric polynomials, *J. Reine Angew. Math.* **283/284** (1976) 265–274
- [19] Srivastava H M and Singh N P, The integration of certain products of the multivariable  $H$ -function with a general class of polynomials, *Rend. Circ. Mat. Palermo (2)* **32** (1983) 157–187



# On the absolute matrix summability of Fourier series and some associated series

B K RAY and A K SAHOO\*

Department of Mathematics, BJB Morning College, Bhubaneswar 751 014, India

\*Department of Mathematics, Kolasib College, Mizoram 796 081, India

MS received 29 January 1993

**Abstract.** The object of the paper is to study the absolute matrix summability problem of Fourier series, conjugate series and some associated series under a new set of conditions on matrix methods, generalising many known results in the literature.

**Keywords.** Absolute summability; triangular matrix; Cesàro mean; Nörlund mean.

## 1. Preliminaries

1.1. Let  $\sum_{n=0}^{\infty} u_n$  be a given infinite series with sequence of partial sums  $\{S_n\}$ . Let  $A = (a_{n,k})$  be an infinite matrix with real or complex elements. The sequence-to-sequence transform

$$t_n = \sum_{k=0}^{\infty} a_{n,k} S_k \quad (1)$$

defines the  $A$ -transform of the sequence  $\{S_n\}$  provided each  $t_n$  exists. If

$$\lim_{n \rightarrow \infty} t_n = S \quad (2)$$

then the series  $\sum_{n=0}^{\infty} u_n$  or the sequence  $\{S_n\}$  is said to be summable  $(A)$ . The series

$\sum_{n=0}^{\infty} u_n$  or the sequence  $\{S_n\}$  is said to be absolutely summable  $A$  or summable  $|A|$ , if  $\{t_n\}$  is of bounded variation, i.e.

$$\sum_{n=1}^{\infty} |t_n - t_{n-1}| < \infty. \quad (3)$$

The method  $(A)$  is said to be conservative or absolutely conservative according as

$$\{S_n\} \in (C, 0) \Rightarrow \{t_n\} \in (C, 0)$$

or

$$\{S_n\} \in |C, 0| \Rightarrow \{t_n\} \in |C, 0|.$$

conservative) and further limit preserving. It is known [8] that  $A$ -method is absolutely regular if and only if

$$(i) \sum_{k=0}^{\infty} a_{n,k} \text{ converges for all } n \geq 0$$

$$(ii) \sum_{n=0}^{\infty} \sum_{k=0}^p |a_{n,k} - a_{n-1,k}| \leq B$$

for every  $p$ , where  $B$  is an absolute constant.

The matrix is called a triangular matrix, if  $a_{n,k} = 0$  for every  $k > n$ . The  $A$ -method reduces to familiar Nörlund method  $(N, p_n)$ , if

$$a_{n,k} = \begin{cases} \frac{p_{n-k}}{P_n}, & k \leq n \\ 0, & k > n \end{cases}$$

where  $\{p_n\}$  is a sequence of constants real or complex such that  $P_n = p_0 + p_1 + \dots$ ,  $p_n \neq 0$ ,  $n \geq 0$ . Important special cases of Nörlund means are Riesz's harmonic mean when  $p_n = 1/(n+1)$  and the Cesàro mean when

$$p_n = \binom{n+\alpha-1}{\alpha-1}, \quad \alpha > 0.$$

1.2. Let  $f(t)$  be a periodic function with period  $2\pi$  and Lebesgue integrable in  $(-\pi, \pi)$ . Let

$$f(t) \sim \frac{1}{2}a_0 + \sum_{n=1}^{\infty} (a_n \cos nt + b_n \sin nt) \equiv \sum_{n=0}^{\infty} A_n(t).$$

The series conjugate to (6) at  $t = x$  is

$$\sum_{n=1}^{\infty} (b_n \cos nx - a_n \sin nx) \equiv \sum_{n=1}^{\infty} B_n(x).$$

We write

$$\Phi(t) = \frac{1}{2} \{f(x+t) + f(x-t) - 2S\},$$

where  $S$  is a function of  $x$ .

$$\psi(t) = \frac{1}{2} \{f(x+t) - f(x-t)\}.$$

If the  $n$ th partial sum of (6) at  $t = x$  is denoted by  $S_n(x)$ , then

$$S_n(x) - S = \frac{2}{\pi} \int_0^{\pi} \Phi(t) \frac{\sin(n + \frac{1}{2})t}{2 \sin \frac{1}{2}t} dt$$

1.3. The absolute Cesàro summability of Fourier series was first studied by Bosanquet and his result reads as follows:

**Theorem A.** (Bosanquet [1]).

$$\Phi(t) \in BV(0, \pi) \Rightarrow \Sigma A_n(x) \in |C, \alpha|, \quad \alpha > 0.$$

Subsequently Bosanquet and Hyslop [2] proved the following theorem for conjugate series.

**Theorem B.** (Bosanquet and Hyslop [2]).

$$\psi(t) \in BV(0, \pi) \text{ and } \frac{\psi(t)}{t} \in L(0, \pi) \Rightarrow \Sigma B_n(x) \in |C, \alpha|, \quad \alpha > 0.$$

The absolute Cesàro summability of  $\Sigma n^\alpha A_n(x)$  and  $\Sigma n^\alpha B_n(x)$ ,  $0 < \alpha < 1$  was studied by Mohanty. He proved the following theorems.

**Theorem C.** (Mohanty [9]). For  $0 < \alpha < 1$

$$\int_0^\pi t^{-\alpha} |d\Phi(t)| < \infty \Rightarrow \Sigma n^\alpha A_n(x) \in |C, \beta|, \quad \beta > \alpha$$

**Theorem D.** (Mohanty [9]). For  $0 < \alpha < 1$

$$\psi(+0) = 0 \text{ and } \int_0^\pi t^{-\alpha} |d\psi(t)| < \infty \Rightarrow \Sigma n^\alpha B_n(x) \in |C, \beta|, \quad \beta > \alpha$$

Mazhar proved the following

**Theorem E.** (Mazhar [7]).

$$\frac{\Phi(t)}{t} \in L(0, \pi) \Rightarrow \Sigma \frac{S_n(x) - S}{n} \in |C, \alpha|, \quad \alpha > 0.$$

In 1971 Nanda Kishore and Hota [5] studied the absolute matrix summability of Fourier series by triangular matrices. Kuttner and Sahaney [6] studied the absolute matrix summability of Fourier series by rectangular matrices. Recently Varshney [12] has improved upon a theorem of Nanda Kishore and Hota [5] by relaxing conditions on the matrix element  $a_{n,k}$ . The absolute matrix summability of a conjugate series has been studied by Hota [3], Ray and Mishra [11] and Jena and Padhy [4].

## 2. Purpose of the present work

2.1. The main object of this paper is to study the absolute matrix summability problem of the series

$$\Sigma A_n(x), \Sigma B_n(x), \Sigma n^\alpha A_n(x), \Sigma n^\alpha B_n(x) \quad (0 < \alpha < 1)$$

and

$$\sum \frac{S_n(x) - S}{n}$$

under a new set of conditions on the matrix element  $a_{n,k}$ . At this stage, we wish to acknowledge that our work has been greatly influenced by a recent paper of Varshney [12] on the absolute matrix summability of Fourier series.

We write

$$A_{n,k} = \sum_{v=k}^n a_{n,v} \neq 0 \quad \text{for all } n \geq 0$$

$$v_{n,k} = (n-k+1) \frac{a_{n,k}}{A_{n,k}}$$

$$\Delta_k p_k = p_k - p_{k+1}.$$

Throughout the present paper  $C$  denotes a positive constant which is not necessarily the same at each occurrence.

We prove the following theorems:

**Theorem 1.** Let  $A = (a_{n,k})$  be an infinite triangular matrix. Suppose that

$$(n+1)(a_{n-1,k} - a_{n,k+1}) \geq \Delta_k \{(n-k)a_{n,k}\} \quad (9)$$

$$a_{n,k} \geq 0 \text{ and } A_{n,0} = 1 \text{ for all } n \text{ and } k \quad (10)$$

$$v_{n,k} \leq C \text{ uniformly for } 0 \leq k \leq n \quad (11)$$

$$\sum_{n=k}^{\infty} \frac{A_{n,n-k}}{n+1} \leq C \text{ for every } k \quad (12)$$

and

$$\sum_{n=r+1}^{\infty} \frac{1}{n+1} \sum_{k=0}^{n-1-r} |\Delta_k(a_{n,k})| \leq \frac{C}{r+1}. \quad (13)$$

Then

$$\Phi(t) \in BV(0, \pi) \Rightarrow \Sigma A_n(x) \in |A|.$$

**Theorem 2.** Let  $A = (a_{n,k})$  be an infinite triangular matrix so that conditions (9)–(13) of Theorem 1 hold. Then

$$\psi(t) \in BV(0, \pi) \text{ and } \frac{\psi(t)}{t} \in L(0, \pi) \Rightarrow \Sigma B_n(x) \in |A|.$$

**Theorem 3.** Let  $A = (a_{n,k})$  be an infinite triangular matrix. If  $0 < \alpha < 1$  and

$$\int_0^{\pi} t^{-\alpha} |d\Phi(t)| < \infty \quad (14)$$

then  $\Sigma n^{\alpha} A_n(x) \in |A|$  provided

$$\sup_k \frac{1}{k^{\alpha}} \sum_{n=k}^{\infty} \frac{A_{n,n-k}}{n^{1-\alpha}} < \infty \quad (15)$$



$$\sum_{n=r+1}^{\infty} \frac{1}{(n+1)^{1-\alpha}} \sum_{k=0}^{n-1-r} |\Delta_k(a_{n,k})| \leq \frac{C}{r^{1-\alpha}} \quad (16)$$

and conditions (9), (10) and (11) of Theorem 1 are satisfied.

**Theorem 4.** Let  $A = (a_{n,k})$  be an infinite triangular matrix. If for  $0 < \alpha < 1$

$$\psi(+0) = 0 \text{ and } \int_0^{\pi} t^{-\alpha} |\mathrm{d}\psi(t)| < \infty \quad (17)$$

then  $\Sigma n^{\alpha} B_n(x) \in |A|$  provided the conditions (9), (10) and (11) of Theorem 1 and (15) and (16) of Theorem 3 are satisfied.

**Theorem 5.** Let  $A = (a_{n,k})$  be an infinite triangular matrix so that conditions (9)–(13) of Theorem 1 hold. Then

$$\frac{\Phi(t)}{t} \in L(0, \pi) \Rightarrow \Sigma \frac{S_n(x) - S}{n} \in |A|.$$

2.2. Remark. Theorems A, B and E follow from Theorems 1, 2 and 5 respectively, if we take

$$a_{n,k} = \begin{cases} \binom{n-k+\alpha-1}{\alpha-1} / \binom{n+\alpha}{\alpha}, \alpha > 0 & \text{for } k \leq n \\ 0 & \text{for } k > n. \end{cases}$$

Further if we take

$$a_{n,k} = \begin{cases} \binom{n-k+\beta-1}{\beta-1} / \binom{n+\beta}{\beta}, 1 > \beta > \alpha > 0 & \text{for } k \leq n \\ 0, & \text{for } k > n \end{cases}$$

Then Theorems C and D follow from Theorems 3 and 4 respectively.

### 3. Notations and Lemmas

#### 3.1. Notation

$$\tau = \left[ \frac{1}{t} \right]$$

$$g(n, t) = \sum_{k=1}^n \frac{k+1}{k} a_{n,k} \sin kt$$

$$h(n, t) = \sum_{k=1}^n \frac{k+1}{k} a_{n,k} \cos kt$$

$$p(n, t) = \sum_{k=1}^n (k+1) a_{n,k} \sin kt$$

$$E(n, t) = \int_t^\pi \frac{p(n, u)}{u} du$$

$$G(n, t) = \sum_{k=1}^n \frac{k+1}{k^{1-\alpha}} a_{n,k} \sin kt, \quad 0 < \alpha < 1$$

$$H(n, t) = \sum_{k=1}^n \frac{k+1}{k^{1-\alpha}} a_{n,k} \cos kt, \quad 0 < \alpha < 1$$

$$L(n, t) = \sum_{k=1}^n \frac{k+1}{k} a_{n,k} \sin\left(k + \frac{1}{2}\right)t.$$

Clearly

$$h'(n, t) = -p(n, t).$$

3.2. We require the following lemmas for proof of our theorems.

*Lemma 1.* (Mohanty and Ray [10]). *If  $\eta > 0$ ,  $\rho > 0$  and  $t^\rho h(t) = H(t)$  then necessary and sufficient conditions that (i)  $h(t) \in BV(0, \eta)$  and  $h(t)/t \in L(0, \eta)$  are that*

$$\int_0^\eta t^{-\rho} |dH(t)| < \infty \text{ and } H(+0) = 0.$$

*Lemma 2.* *Suppose the matrix element  $a_{n,k}$  satisfies conditions (9), (10) and (11) of Theorem 1. Further suppose that*

$$S_n = \sum_{k=0}^n u_k = O(n^\alpha) \text{ for } 0 < \alpha < 1 \quad (18)$$

*Then  $\sum u_k \in |A|$  if and only if*

$$\sum_{n=1}^{\infty} \frac{1}{n+1} \left| \sum_{k=0}^n (k+1) a_{n,k} u_k \right| < \infty. \quad (19)$$

This is an improved version of an earlier lemma due to Jena and Padhy ([4], Lemma 3).

*Proof.* Let

$$\lambda_{n,k} = (n+1)(A_{n,k} - A_{n-1,k}) - (k+1)a_{n,k} \quad (20)$$

By simple calculation, we obtain

$$\begin{aligned} \lambda_{n,k} &= (A_{n,k} - A_{n-1,k} - a_{n,k})(n+1) + (n-k)a_{n,k} \\ \lambda_{n,n} &= 0, \quad \lambda_{n,0} = -a_{n,0} \\ \Delta_k(\lambda_{n,k}) &= \Delta_k\{(n-k)a_{n,k}\} - (n+1)(a_{n-1,k} - a_{n,k+1}) \leq 0 \end{aligned} \quad (21)$$

by condition (9). Since the matrix  $(a_{n,k})$  is triangular, we have

$$t_n = \sum_{k=0}^n a_{n,k} S_k = \sum_{k=0}^n A_{n,k} u_k,$$

where  $t_n$  is the  $A$ -transform of  $\{S_n\}$ .

$$\begin{aligned}
&= \frac{1}{n+1} \sum_{k=0}^n (\lambda_{n,k} + (k+1)a_{n,k})u_k, \quad \text{using (20)} \\
&= \frac{1}{n+1} \sum_{k=0}^n \lambda_{n,k}u_k + \frac{1}{n+1} \sum_{k=0}^n (k+1)a_{n,k}u_k \\
&= X_n + Y_n, \quad \text{say.} \tag{22}
\end{aligned}$$

Since  $\lambda_{n,n} = 0$  and  $\lambda_{n,0} = -a_{n,0}$ , we obtain by Abel's method of partial summation

$$X_n = \frac{1}{n+1} \sum_{k=0}^{n-1} \Delta_k(\lambda_{n,k})S_k.$$

Using (18), we get

$$\begin{aligned}
|X_n| &\leq \frac{C}{n+1} \sum_{k=0}^{n-1} |\Delta_k(\lambda_{n,k})|k^\alpha \\
&\leq Cn^{\alpha-1} \sum_{k=0}^{n-1} |\Delta_k(\lambda_{n,k})| \\
&= -Cn^{\alpha-1} \sum_{k=0}^{n-1} \Delta_k(\lambda_{n,k}), \quad \text{using (21)} \\
&= -Cn^{\alpha-1}(\lambda_{n,0} - \lambda_{n,n}) \\
&= Cn^{\alpha-1}a_{n,0} = O(n^{\alpha-2})
\end{aligned}$$

as

$$a_{n,0} = \frac{v_{n,0}}{n+1} \leq \frac{C}{n+1} \quad \text{by (11).}$$

So

$$\sum_{n=1}^{\infty} |X_n| = O\left(\sum_{n=1}^{\infty} \frac{1}{n^{2-\alpha}}\right) = O(1), \quad 0 < \alpha < 1 \tag{23}$$

Now, the lemma follows at once from (22) and (23).

*Lemma 3.*

- (i) [13]  $\Phi(t) \in BV(0, \pi) \Rightarrow \sum A_n(x) \in (C, 0)$
- (ii) [13]  $\frac{\psi(t)}{t} \in L(0, \pi) \Rightarrow \sum B_n(x) \in (C, 0)$
- (iii)  $\Phi(t) \in BV(0, \pi) \Rightarrow \sum_{k=1}^n k^\alpha A_k(x) = O(n^\alpha), \quad 0 < \alpha < 1$
- (iv)  $\psi(t) \in BV(0, \pi) \Rightarrow \sum_{k=1}^n k^\alpha B_k(x) = O(n^\alpha), \quad 0 < \alpha < 1$
- (v)  $\frac{\Phi(t)}{t} \in L(0, \pi) \Rightarrow \sum \frac{S_n(x) - S}{n} \in (C, 0)$

*Proof.* (iii) and (iv) follow at once, since

$$\Phi(t) \in BV(0, \pi) \Rightarrow A_n(x) = O(n^{-1})$$

and

$$\psi(t) \in BV(0, \pi) \Rightarrow B_n(x) = O(n^{-1}).$$

It is known that the condition  $\Phi(t)/t \in L(0, \pi)$  implies

$$\int_0^t |\Phi(u)| du = o(t) \text{ and } \int_0^\pi \frac{\Phi(u)}{u} du \text{ exists.} \quad (24)$$

Part (v) follows at once as  $\Sigma[S_n(x) - S]/n$  is known ([13], p. 125) to be convergent under condition (24).

*Lemma 4.*

$$(i) \quad g(n, t) = O(nt)$$

$$(ii) \quad g(n, t) = O\left\{t^{-1} \sum_{k=1}^{n-\tau-2} |\Delta_k(a_{n,k})|\right\} + O(A_{n,n-\tau-1}), n > \tau$$

$$(iii) \quad h(n, t) = O\left\{t^{-1} \sum_{k=1}^{n-\tau-2} |\Delta_k(a_{n,k})|\right\} + O(A_{n,n-\tau-1}), n > \tau$$

$$(iv) \quad E(n, t) = O(n)$$

$$(v) \quad E(n, t) = O\left\{t^{-2} \sum_{k=1}^{n-\tau-2} |\Delta_k(a_{n,k})|\right\} + O\left\{t^{-1} A_{n,n-\tau-1}\right\}, n > \tau$$

*Proof of (i).* We have

$$\begin{aligned} g(n, t) &= \sum_{k=1}^n \frac{k+1}{k} a_{n,k} \sin kt \\ &= O\left(t \sum_{k=1}^n (k+1) a_{n,k}\right) \\ &= O\left(nt \sum_{k=1}^n a_{n,k}\right) = O(nt). \end{aligned}$$

*Proof of (ii).* For  $n > \tau$

$$\begin{aligned} g(n, t) &= \sum_{k=1}^{n-\tau-1} \frac{k+1}{k} a_{n,k} \sin kt + \sum_{k=n-\tau}^n \frac{k+1}{k} a_{n,k} \sin kt \\ &= \sum_{k=1}^{n-\tau-2} \left( \sum_{r=1}^k \frac{r+1}{r} \sin rt \right) \Delta_k(a_{n,k}) \\ &\quad + a_{n,n-\tau-1} \left( \sum_{r=1}^{n-\tau-1} \frac{r+1}{r} \sin rt \right) + \sum_{k=n-\tau}^n \frac{k+1}{k} a_{n,k} \sin kt. \end{aligned}$$

Now

$$\begin{aligned}
 |g(n, t)| &\leq C \left[ \frac{1}{t} \sum_{k=1}^{n-\tau-2} |\Delta_k(a_{n,k})| + \frac{1}{t} a_{n,n-\tau-1} + A_{n,n-\tau} \right] \\
 &= 0 \left\{ t^{-1} \sum_{k=1}^{n-\tau-2} |\Delta_k(a_{n,k})| \right\} + 0 \left\{ \frac{A_{n,n-\tau-1}}{t(\tau+2)} \right\} + 0 \left\{ A_{n,n-\tau} \right\},
 \end{aligned}$$

since

$$a_{n,n-\tau-1} = \frac{A_{n,n-\tau-1} V_{n,n-\tau-1}}{\tau+2} \leq \frac{CA_{n,n-\tau-1}}{\tau+2}.$$

Thus, we get

$$g(n, t) = 0 \left\{ t^{-1} \sum_{k=1}^{n-\tau-2} |\Delta_k(a_{n,k})| \right\} + 0 \left\{ A_{n,n-\tau-1} \right\} \quad (25)$$

Adopting the technique used above it can be proved that for  $t < t' < \pi$  (replacing  $\sin kt$  by  $\sin kt'$  but retaining splitting of the sum in tact as in the case of  $g(n, t)$ ), it can be proved that

$$\begin{aligned}
 g(n, t') &= 0 \left\{ t'^{-1} \sum_{k=1}^{n-\tau-2} |\Delta_k(a_{n,k})| \right\} + 0 \left\{ A_{n,n-\tau-1} \right\} \\
 &= 0 \left\{ t^{-1} \sum_{k=1}^{n-\tau-2} |\Delta_k(a_{n,k})| \right\} + 0 \left\{ A_{n,n-\tau-1} \right\}
 \end{aligned} \quad (26)$$

We omit the proof of (iii) as it is same (replace  $\sin kt$  by  $\cos kt$ ) as that of (ii). At this stage, we remark that for  $t < t' < \pi$ ,  $h(n, t')$  has the same estimate as that of  $h(n, t)$ .

*Proof of (iv).* We have

$$\begin{aligned}
 E(n, t) &= \int_t^\pi \frac{p(n, u)}{u} du \\
 &= \sum_{k=1}^n (k+1) a_{n,k} \int_t^\pi \frac{\sin ku}{u} du \\
 &= 0 \left( \sum_{k=1}^n (k+1) a_{n,k} \right) = 0(n)
 \end{aligned}$$

*Proof of (v).* By mean value theorem, we have for  $t < t' < \pi$

$$\begin{aligned}
 E(n, t) &= \sum_{k=1}^n (k+1) a_{n,k} \int_t^\pi \frac{\sin ku}{u} du \\
 &= t^{-1} \sum_{k=1}^n \frac{k+1}{k} a_{n,k} (\cos kt - \cos kt') \\
 &= t^{-1} \{ h(n, t) - h(n, t') \} \\
 &= 0 \left\{ t^{-2} \sum_{k=1}^{n-\tau-2} |\Delta_k(a_{n,k})| \right\} + 0 \left\{ t^{-1} A_{n,n-\tau-1} \right\},
 \end{aligned}$$

using (iii) and the fact that  $h(n, t')$  has the same estimate as that of  $h(n, t)$ .

- (i)  $G(n, t) = 0(n^{1+\alpha}t)$
- (ii)  $G(n, t) = 0\left\{t^{-1}n^\alpha \sum_{k=1}^{n-\tau-2} |\Delta_k(a_{n,k})|\right\} + 0\left\{n^\alpha A_{n,n-\tau-1}\right\}, n > \tau$
- (iii)  $H(n, t) = 0(n^\alpha)$
- (iv)  $H(n, t) = 0\left\{t^{-1}n^\alpha \sum_{k=1}^{n-\tau-2} |\Delta_k(a_{n,k})|\right\} + 0\{n^\alpha A_{n,n-\tau-1}\}, n > \tau$
- (v)  $L(n, t) = 0(nt)$
- (vi)  $L(n, t) = 0\left\{t^{-1} \sum_{k=1}^{n-\tau-2} |\Delta_k(a_{n,k})|\right\} + 0(A_{n,n-\tau-1}), n > \tau$

(i), (iii) and (v) can be proved adopting the technique employed in proving Lemma 4(i). Proof of (vi) is same as the proof of Lemma 4(ii).

*Proof of (ii).* For  $n > \tau$

$$\begin{aligned}
 G(n, t) &= \sum_{k=1}^{n-\tau-1} \frac{k+1}{k^{1-\alpha}} a_{n,k} \sin kt + \sum_{k=n-\tau}^n \frac{k+1}{k^{1-\alpha}} a_{n,k} \sin kt \\
 &= \sum_{k=1}^{n-\tau-2} \left( \sum_{r=1}^k \frac{r+1}{r^{1-\alpha}} \sin rt \right) \Delta_k(a_{n,k}) \\
 &\quad + a_{n,n-\tau-1} \left( \sum_{r=1}^{n-\tau-1} \frac{r+1}{r^{1-\alpha}} \sin rt \right) + \sum_{k=n-\tau}^n \frac{k+1}{k^{1-\alpha}} a_{n,k} \sin kt \\
 &= 0\left\{n^\alpha t^{-1} \sum_{k=1}^{n-\tau-2} |\Delta_k(a_{n,k})|\right\} + 0\{n^\alpha t^{-1} a_{n,n-\tau-1}\} + 0\{n^\alpha A_{n,n-\tau}\} \\
 &= 0\left\{t^{-1} n^\alpha \sum_{k=1}^{n-\tau-2} |\Delta_k(a_{n,k})|\right\} + 0\{n^\alpha A_{n,n-\tau-1}\},
 \end{aligned}$$

since

$$\frac{a_{n,n-\tau-1}}{t} = \frac{A_{n,n-\tau-1}}{t} \frac{V_{n,n-\tau-1}}{(\tau+2)} \leq C A_{n,n-\tau-1}.$$

Proof of (iv) is similar to the proof of (ii).

#### 4. Proof of the theorems

4.1. *Proof of theorem 1.* By Lemma 3(i) the  $n$ th partial sum of  $\Sigma A_n(x)$  remains bounded under the hypothesis  $\Phi(t) \in BV(0, \pi)$  and hence by Lemma 2 the series  $\Sigma A_n(x) \in |A|$ , if and only if

$$I = \sum_{n=1}^{\infty} \frac{1}{n+1} \left| \sum_{k=0}^n (k+1) a_{n,k} A_k(x) \right| < \infty. \quad (27)$$

For  $k \geq 1$

$$\frac{\pi}{2} A_k(x) = \int_0^\pi \Phi(t) \cos kt \, dt = - \int_0^\pi \frac{\sin kt}{k} d\Phi(t).$$

Putting this value of  $A_k(x)$  in (27) and using the notations given in § 3.1, we find that

$$\begin{aligned} |I| &\leq \frac{2}{\pi} \int_0^\pi |d\Phi(t)| \sum_{n=1}^\infty \frac{1}{n+1} \left| \sum_{k=1}^n \frac{k+1}{k} a_{n,k} \sin kt \right| \\ &= \frac{2}{\pi} \int_0^\pi |d\Phi(t)| \sum_{n=1}^\infty \frac{1}{n+1} |g(n, t)|. \end{aligned}$$

Since  $\int_0^\pi |d\Phi(t)|$  is finite it is enough to show that uniformly in  $0 < t < \pi$

$$\sum_{n=1}^\infty \frac{1}{n+1} |g(n, t)| = O(1). \quad (28)$$

Applying (i) and (ii) of Lemma 4, we get

$$\begin{aligned} \sum_{n=1}^\infty \frac{|g(n, t)|}{n+1} &= \sum_{n \leq \tau} \frac{|g(n, t)|}{n+1} + \sum_{n > \tau} \frac{|g(n, t)|}{n+1} \\ &= O\left(\sum_{n \leq \tau} t\right) + O\left\{t^{-1} \sum_{n > \tau} \frac{1}{n+1} \sum_{k=1}^{n-\tau-2} |\Delta_k(a_{n,k})|\right\} \\ &\quad + O\left(\sum_{n > \tau} \frac{A_{n, n-\tau-1}}{n+1}\right) \\ &= O(1) + O\frac{1}{t(\tau+1)} + O(1), \text{ using (12) and (13)} \\ &= O(1). \end{aligned}$$

This completes the proof of Theorem 1.

4.2. *Proof of theorem 2.* By Lemma 3(ii) the  $n$ th partial sum of  $\Sigma B_n(x)$  remains bounded whenever  $\psi(t)/t \in L(0, \pi)$  and hence by Lemma 2 the series  $\Sigma B_n(x) \in |A|$ , if and only if

$$J = \sum_{n=1}^\infty \frac{1}{n+1} \left| \sum_{k=1}^n (k+1) a_{n,k} B_k(x) \right| < \infty \quad (29)$$

Writing

$$\alpha(k, t) = \int_t^\pi \frac{\sin ku}{u} du, \text{ we have for } k \geq 1$$

$$\frac{\pi}{2} B_k(x) = \int_0^\pi \psi(t) \sin kt \, dt$$

$$\begin{aligned}
&= \int_0^\pi d\{t\psi(t)\} \alpha(k, t) dt, \\
&= \int_0^\pi d\{t\psi(t)\} \int_t^\pi \frac{\sin ku}{u} du,
\end{aligned} \tag{30}$$

since  $\alpha(k, \pi) = 0$ ,  $\psi(+0)$  and  $\alpha(k, 0)$  are finite.

Putting this value of  $B_k(x)$  in (29) and using the notations given in § 3.1, we have

$$\begin{aligned}
|J| &\leq \frac{2}{\pi} \int_0^\pi |d(t\psi(t))| \left| \sum_{n=1}^\infty \frac{1}{n+1} \right| \int_t^\pi \frac{1}{u} \sum_{k=1}^n (k+1) a_{n,k} \sin ku du \\
&= \frac{2}{\pi} \int_0^\pi |d(t\psi(t))| \sum_{n=1}^\infty \frac{|E(n, t)|}{n+1}.
\end{aligned}$$

The integral  $\int_0^\pi t^{-1} |d(t\psi(t))|$  is finite by Lemma 1 and hence it is enough to show that uniformly in  $0 < t < \pi$

$$\sum_{n=1}^\infty \frac{|E(n, t)|}{n+1} = O(t^{-1}). \tag{31}$$

Now applying (iv) and (v) of Lemma 4, we get

$$\begin{aligned}
\sum_{n=1}^\infty \frac{|E(n, t)|}{n+1} &= \sum_{n \leq \tau} \frac{|E(n, t)|}{n+1} + \sum_{n > \tau} \frac{|E(n, t)|}{n+1} \\
&= O\left(\sum_{n \leq \tau} \frac{n}{n+1}\right) + O\left(t^{-2} \sum_{n > \tau} \frac{1}{n+1} \sum_{k=1}^{n-\tau-2} |\Delta_k(a_{n,k})|\right) \\
&\quad + O\left(t^{-1} \sum_{n > \tau} \frac{A_{n, n-\tau-1}}{n+1}\right) \\
&= O(t^{-1}) + O\left(\frac{t^{-2}}{\tau+1}\right) + O(t^{-1}) = O(t^{-1}),
\end{aligned}$$

using (12) and (13).

This completes the proof of Theorem 2.

4.3. *Proof of theorem 3.* By Lemma 3(iii), we get

$$\int_0^\pi t^{-\alpha} |d\Phi(t)| < \infty \Rightarrow \sum_{k=1}^n k^\alpha A_k(x) = O(n^\alpha)$$

and hence by Lemma 2 the series  $\sum n^\alpha A_n(x) \in |A|$  if and only if

$$K = \sum_{n=1}^\infty \frac{1}{n+1} \left| \sum_{k=1}^n (k+1) a_{n,k} k^\alpha A_k(x) \right| < \infty. \tag{32}$$



Putting  $A_k(x) = -\frac{2}{\pi} \int_0^\pi \frac{\sin kt}{t} d\Phi(t)$  in (32) and using the notations of § 3.1, we obtain

$$\begin{aligned} |K| &\leq \frac{2}{\pi} \int_0^\pi |d\Phi(t)| \sum_{n=1}^\infty \frac{1}{n+1} \left| \sum_{k=1}^n \frac{k+1}{k^{1-\alpha}} a_{n,k} \sin kt \right| \\ &= \frac{2}{\pi} \int_0^\pi |d\Phi(t)| \sum_{n=1}^\infty \frac{|G(n, t)|}{n+1}. \end{aligned}$$

By the hypothesis  $\int_0^\pi t^{-\alpha} |d\Phi(t)|$  is finite and hence it suffices to show that uniformly in  $0 < t \leq \pi$

$$\sum \equiv \sum_{n=1}^\infty \frac{|G(n, t)|}{n+1} = O(t^{-\alpha}). \quad (33)$$

Writing  $\sum = \sum_{n \leq \tau} + \sum_{n > \tau}$  and using Lemma 5(i) and Lemma 5(ii), we get

$$\begin{aligned} \sum &= O\left(t \sum_{n \leq \tau} \frac{n^{1+\alpha}}{n+1}\right) + O\left\{t^{-1} \sum_{n > \tau} \frac{n^\alpha}{n+1} \sum_{k=1}^{n-\tau-2} |\Delta_k(a_{n,k})|\right\} \\ &\quad + O\left\{\sum_{n > \tau} \frac{n^\alpha}{n+1} A_{n, n-\tau-1}\right\} \\ &= O(t\tau^{1+\alpha}) + O\left\{t^{-1} \sum_{n > \tau} \frac{1}{n^{1-\alpha}} \sum_{k=1}^{n-\tau-2} |\Delta_k(a_{n,k})|\right\} + O\left\{\sum_{n > \tau} \frac{A_{n, n-\tau-1}}{n^{1-\alpha}}\right\} \\ &= O(t^{-\alpha}) + O\left\{\frac{1}{t\tau^{1-\alpha}}\right\} + O\{\tau^\alpha\}, \text{ using (15) and (16)} \\ &= O(t^{-\alpha}). \end{aligned}$$

This completes the proof of Theorem 3.

4.4. *Proof of theorem 4.* By Lemma 3(iv) the  $n$ th partial sum of the series  $\sum n^\alpha B_n(x)$  is  $O(n^\alpha)$  under the assumption  $\int_0^\pi t^{-\alpha} |d\psi(t)| < \infty$  and hence by virtue of Lemma 2 the series  $\sum n^\alpha B_n(x) \in |A|$  if and only if

$$L = \sum_{n=1}^\infty \frac{1}{n+1} \left| \sum_{k=1}^n (k+1) a_{n,k} k^\alpha B_k(x) \right| < \infty \quad (34)$$

Since  $\psi(+0) = 0$ , we have for  $k \geq 1$

$$\begin{aligned} \frac{\pi}{2} B_k(x) &= \int_0^\pi \psi(t) \sin kt \, dt \\ &= -\psi(\pi) \frac{\cos k\pi}{k} + \int_0^\pi \frac{\cos kt}{k} d\psi(t). \end{aligned}$$

Putting the value of  $B_k(x)$  in (34) and using the notations of § 3.1, we get

$$\begin{aligned}
 |L| &\leq \frac{2}{\pi} |\psi(\pi)| \sum_{n=1}^{\infty} \frac{1}{n+1} \left| \sum_{k=1}^n \frac{k+1}{k^{1-\alpha}} a_{n,k} \cos k\pi \right| \\
 &\quad + \frac{2}{\pi} \int_0^{\pi} |d\psi(t)| \sum_{n=1}^{\infty} \frac{1}{n+1} \left| \sum_{k=1}^n \frac{k+1}{k^{1-\alpha}} a_{n,k} \cos kt \right| \\
 &= \frac{2}{\pi} |\psi(\pi)| \sum_{n=1}^{\infty} \frac{|H(n, \pi)|}{n+1} + \frac{2}{\pi} \int_0^{\pi} |d\psi(t)| \sum_{n=1}^{\infty} \frac{|H(n, t)|}{n+1} \\
 &= L_1 + L_2, \text{ say}
 \end{aligned} \tag{35}$$

Writing  $\sum_{n=1}^{\infty} \frac{H(n, t)}{n+1} \equiv \sum_{n \leq \tau} + \sum_{n > \tau}$  and using (iii) and (iv) of Lemma 5, we can prove (similar to the proof of (33)) that

$$\sum_{n=1}^{\infty} \frac{|H(n, t)|}{n+1} = O(t^{-\alpha}) \text{ uniformly in } 0 < t \leq \pi$$

and hence

$$L_2 = O\left(\int_0^{\pi} t^{-\alpha} |d\psi(t)|\right) = O(1). \tag{36}$$

We have

$$\begin{aligned}
 H(n, \pi) &= \sum_{k=1}^n \frac{k+1}{k^{1-\alpha}} a_{n,k} \cos k\pi \\
 &= \sum_{k=1}^{n-1} \left( \sum_{r=1}^k \frac{r+1}{r^{1-\alpha}} \cos r\pi \right) \Delta_k(a_{n,k}) + a_{n,n} \sum_{r=1}^n \frac{r+1}{r^{1-\alpha}} \cos k\pi \\
 &= O\left\{ \sum_{k=1}^{n-1} |\Delta_k(a_{n,k})| k^{\alpha} \right\} + O(n^{\alpha} a_{n,n}) \\
 &= O\left\{ n^{\alpha} \sum_{k=1}^{n-1} |\Delta_k(a_{n,k})| \right\} + O(n^{\alpha} A_{n,n-1}).
 \end{aligned}$$

Using this estimate for  $H(n, \pi)$ , we get

$$\begin{aligned}
 L_1 &= O\left( \sum_{n=1}^{\infty} \frac{1}{n^{1-\alpha}} \sum_{k=1}^{n-1} |\Delta_k(a_{n,k})| \right) + O\left( \sum_{n=1}^{\infty} \frac{A_{n,n-1}}{n^{1-\alpha}} \right) \\
 &= O(1) \text{ by (15) and (16).}
 \end{aligned} \tag{37}$$

Collecting the results of (35), (36) and (37), we get

$$L = O(1),$$

and this completes the proof of theorem 4.

4.5. *Proof of theorem 5.* By Lemma 3(v) the  $n$ th partial sum of  $\sum \frac{S_n(x) - S}{n}$  is bounded

whenever  $\frac{\Phi(t)}{t} \in L(0, \pi)$  and hence by Lemma 2 the series  $\sum \frac{S_n(x) - S}{n} \in |A|$ , if and only if

$$M = \sum_{n=1}^{\infty} \frac{1}{n+1} \left| \sum_{k=1}^n (k+1) a_{n,k} \frac{S_k(x) - S}{k} \right| < \infty. \quad (38)$$

For  $k \geq 1$

$$S_k(x) - S = \frac{2}{\pi} \int_0^{\pi} \Phi(t) \frac{\sin(k + \frac{1}{2})t}{2 \sin \frac{1}{2}t} dt.$$

Putting these values of  $S_k(x) - S$  in (38) and using the notation of § 3.1, we get

$$\begin{aligned} M &= \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{1}{n+1} \left| \int_0^{\pi} \frac{\Phi(t)}{2 \sin \frac{1}{2}t} \sum_{k=1}^n \frac{k+1}{k} a_{n,k} \sin\left(k + \frac{1}{2}\right)t dt \right| \\ &\leq \int_0^{\pi} \frac{|\Phi(t)|}{2 \sin \frac{1}{2}t} dt \sum_{n=1}^{\infty} \frac{1}{n+1} \left| \sum_{k=1}^n \frac{k+1}{k} a_{n,k} \sin\left(k + \frac{1}{2}\right)t \right| \\ &= \int_0^{\pi} \frac{|\Phi(t)|}{2 \sin \frac{1}{2}t} dt \sum_{n=1}^{\infty} \frac{|L(n, t)|}{n+1}. \end{aligned} \quad (39)$$

Adopting the technique similar to those used in establishing (28) in the proof of Theorem 1 we can show that uniformly in  $0 < t \leq \pi$

$$\sum_{n=1}^{\infty} \frac{|L(n, t)|}{n+1} = o(1). \quad (40)$$

From (39) and (40), it follows that

$$M = o\left(\int_0^{\pi} \frac{|\Phi(t)|}{2 \sin \frac{1}{2}t} dt\right) = o(1),$$

by the hypothesis and this completes the proof of Theorem 5.

## References

- [1] Bosanquet L S, Note on absolute summability (C) of a Fourier series, *J. London Math. Soc.* **11** (1936) 11-15
- [2] Bosanquet L S and Hyslop J M, On the absolute summability of the allied series of a Fourier series, *Math. Zeit.* **42** (1937) 489-512
- [3] Hota G C, Absolute matrix summability of the conjugate series of a Fourier series, *Aligarh Bull. Math.* Vols 3-4 (1973-1974) 89-98
- [4] Jena S C and Padhy P, Absolute matrix summability of the allied series of a Fourier series, *Proc. Indian Acad. Sci. (Math. Sci.)* **98** (1988) 43-52
- [5] Nanda Kishore and Hota G C, On absolute matrix summability of a Fourier series, *Indian J. Math.* **13** (1971) 99-110
- [6] Kuttner B and Sahanev B N, On the absolute matrix summability of Fourier series, *Pacific J. Math.* **43** (1972) 407-419

- [7] Mazhar S M, On the absolute convergence of a series associated with a Fourier series, *Math. Scandinavica*, **21** (1967) 90–104
- [8] Mears F M, Absolute regularity and the Nörlund mean, *Anal. Math.* **38** (1937) 594–601
- [9] Mohanty R, The absolute Cesàro summability of some series associated with a Fourier series and its allied series, *J. London Math. Soc.* **25** (1950) 63–67
- [10] Mohanty R and Ray B K, On the behaviour of a series associated with the conjugate series of a Fourier series, *Can. J. Math.* **31** (1969) 535–551
- [11] Ray B K and Arati Mishra, On the absolute matrix summability of the series conjugate to a Fourier series, *Indian J. Pure Appl. Math.* **11** (1980) 1482–1496
- [12] Varshney O P, On the absolute matrix summability of Fourier series, *Indian J. Pure Appl. Math.*, **21** (1990) 1015–1023
- [13] Zygmund A, *Trigonometric series* (Vols. I and II combined) (Cambridge: University Press) New York (1968)

## On absolute summability factors of infinite series

HÜSEYİN BOR

Department of Mathematics, Erciyes University, Kayseri 38039, Turkey  
 Mailing address: P K 213, Kayseri 38002, Turkey

MS received 10 April 1993

**Abstract.** In this paper using  $\delta$ -quasi-monotone sequences a theorem on  $|\bar{N}, p_n; \delta|_k$  summability factors of infinite series, which generalizes a theorem of Bor [4] on  $|\bar{N}, p_n|_k$  summability factors of infinite series, is proved. Also, in the special case this theorem includes a result of Mazhar [8] on  $|C, 1|_k$  summability factors.

**Keywords.** Absolute summability; summability factors; infinite series.

### 1. Introduction

A sequence  $(b_n)$  of positive numbers is said to be quasi monotone if  $n\Delta b_n \geq -\alpha b_n$  for some  $\alpha$  and it is said to be  $\delta$ -quasi monotone, if  $b_n \rightarrow 0$ ,  $b_n > 0$  ultimately and  $\Delta b_n \geq -\delta_n$ , where  $(\delta_n)$  is a sequence of positive numbers (see [1]). Let  $\Sigma a_n$  be a given infinite series with partial sums  $(s_n)$ . By  $u_n$  and  $t_n$  we denote the  $n$ th  $(C, 1)$  means of the sequences  $(s_n)$  and  $(na_n)$ , respectively. The series  $\Sigma a_n$  is said to be summable  $|C, 1; \gamma|_k$ ,  $k \geq 1$  and  $\gamma \geq 0$ , if (see [5])

$$\sum_{n=1}^{\infty} n^{\gamma k + k - 1} |u_n - u_{n-1}|^k < \infty. \quad (1.1)$$

But since  $t_n = n(u_n - u_{n-1})$  (see [7]), the condition (1.1) can also be written as

$$\sum_{n=1}^{\infty} n^{\gamma k - 1} |t_n|^k < \infty. \quad (1.2)$$

Let  $(p_n)$  be a sequence of positive numbers such that

$$P_n = \sum_{v=0}^n p_v \rightarrow \infty \quad \text{as } n \rightarrow \infty, \quad (P_{-i} = p_{-i} = 0, i \geq 1). \quad (1.3)$$

The sequence to sequence transformation

$$w_n = \frac{1}{P_n} \sum_{v=0}^n p_v s_v \quad (1.4)$$

defines the sequence  $(w_n)$  of the  $(\bar{N}, p_n)$  means of the sequence  $(s_n)$ , generated by the sequence of coefficients  $(p_n)$  (see [6]). The series  $\Sigma a_n$  is said to be summable  $|\bar{N}, p_n|_k$ ,

$k \geq 1$ , if (see [2])

$$\sum_{n=1}^{\infty} (P_n/p_n)^{k-1} |w_n - w_{n-1}|^k < \infty, \quad (1.5)$$

and it is said to be summable  $|\bar{N}, p_n; \gamma|_k$ ,  $k \geq 1$  and  $\gamma \geq 0$ , if (see [3])

$$\sum_{n=1}^{\infty} (P_n/p_n)^{\gamma k + k - 1} |w_n - w_{n-1}|^k < \infty. \quad (1.6)$$

In the special case when  $p_n = 1$  for all values of  $n$  (resp.  $\gamma = 0$ ),  $|\bar{N}, p_n; \gamma|_k$  summability is the same as  $|C.1; \gamma|_k$  (resp.  $|\bar{N}, p_n|_k$ ) summability. If we write

$$X_n = \sum_{v=0}^n \frac{p_v}{P_v}, \quad (1.7)$$

then  $(X_n)$  is a positive increasing sequence tending to infinity with  $n$ .

## 2.

Quite recently Bor[4] has proved the following theorem for  $|\bar{N}, p_n|_k$  summability factors by using  $\delta$ -quasi monotone sequences.

**Theorem A.** Let  $\lambda_n \rightarrow 0$  as  $n \rightarrow \infty$  and let  $(p_n)$  be a sequence of positive numbers such that

$$P_n = O(np_n) \quad \text{as } n \rightarrow \infty. \quad (2.1)$$

Suppose that there exists a sequence of numbers  $(A_n)$  which is  $\delta$ -quasi monotone with  $\Sigma n \delta_n X_n < \infty$ ,  $\Sigma A_n X_n$  is convergent and  $|\Delta \lambda_n| \leq |A_n|$  for all  $n$ . If

$$\sum_{n=1}^m \frac{p_n}{P_n} |t_n|^k = O(X_m) \quad \text{as } m \rightarrow \infty, \quad (2.2)$$

then the series  $\Sigma a_n \lambda_n$  is summable  $|\bar{N}, p_n|_k$ ,  $k \geq 1$ .

## 3.

The aim of this paper is to prove Theorem A for  $|\bar{N}, p_n; \gamma|_k$  summability. Now, we shall prove the following theorem.

**Theorem.** Let  $\lambda_n \rightarrow 0$  as  $n \rightarrow \infty$  and let  $(p_n)$  be a sequence of positive numbers such that the condition (2.1) is satisfied and

$$\sum_{n=v}^{\infty} (P_n/p_n)^{\gamma k - 1} \frac{1}{P_{n-1}} = O \left\{ (P_v/p_v)^{\gamma k} \frac{1}{P_v} \right\}. \quad (3.1)$$

$\Sigma n \delta_n X_n < \infty$ ,  $\Sigma A_n X_n$  is convergent and  $|\Delta \lambda_n| \leq |A_n|$  for all  $n$ . If

$$\sum_{n=1}^m (P_n/p_n)^{\gamma k-1} |t_n|^k = O(X_m) \quad \text{as } m \rightarrow \infty, \quad (3.2)$$

then the series  $\Sigma a_n \lambda_n$  is summable  $|\bar{N}, p_n; \gamma|_k$  for  $k \geq 1$  and  $\gamma \geq 0$ .

If we take  $\gamma = 0$  in this theorem, then we get Theorem A. Because in this case the condition (3.2) reduces to the condition (2.2). Also in this case the condition (3.1) reduces to

$$\sum_{n=v}^{\infty} \frac{P_n}{P_n P_{n-1}} = O(1/P_v),$$

but this result always holds.

#### 4.

We need the following lemmas for the proof of our theorem.

*Lemma 1* ([4]). *Under the conditions of the theorem, we have*

$$|\lambda_n| X_n = O(1) \quad \text{as } n \rightarrow \infty. \quad (4.1)$$

*Lemma 2* ([4]). *If  $(A_n)$  is  $\delta$ -quasi monotone with  $\Sigma n \delta_n X_n < \infty$  and  $\Sigma A_n X_n$  is convergent, then*

$$m A_m X_m = O(1) \quad \text{as } m \rightarrow \infty \quad (4.2)$$

$$\sum_{n=1}^{\infty} n X_n |\Delta A_n| < \infty. \quad (4.3)$$

#### 5. Proof of the theorem

Let  $(T_n)$  be the sequence of  $(\bar{N}, p_n)$  means of the series  $\Sigma a_n \lambda_n$ . Then, by definition, we have that

$$T_n = \frac{1}{P_n} \sum_{v=0}^n p_v \sum_{i=0}^v a_i \lambda_i = \frac{1}{P_n} \sum_{v=0}^n (P_n - P_{v-1}) a_v \lambda_v. \quad (5.1)$$

Then, for  $n \geq 1$ , we obtain

$$T_n - T_{n-1} = \frac{p_n}{P_n P_{n-1}} \sum_{v=1}^n P_{v-1} a_v \lambda_v = \frac{p_n}{P_n P_{n-1}} \sum_{v=1}^n \frac{P_{v-1} \lambda_v}{v} v a_v. \quad (5.2)$$

Using Abel's transformation, we have

$$T_n - T_{n-1} = \frac{p_n}{P_n P_{n-1}} \sum_{v=1}^{n-1} \Delta \left( \frac{P_{v-1} \lambda_v}{v} \right) \sum_{i=1}^v i a_i + \frac{p_n \lambda_n}{n P_n} \sum_{v=1}^n v a_v$$

$$\begin{aligned}
&= \frac{(P_n/p_n)^{k-1} P_n}{n P_n} - \frac{P_n}{P_n P_{n-1}} \sum_{v=1}^{n-1} \frac{P_v t_v \lambda_v}{v} \\
&\quad + \frac{P_n}{P_n P_{n-1}} \sum_{v=1}^{n-1} P_v \Delta \lambda_v t_v \frac{v+1}{v} + \frac{P_n}{P_n P_{n-1}} \sum_{v=1}^{n-1} P_v \lambda_{v+1} t_v \frac{1}{v} \\
&= T_{n,1} + T_{n,2} + T_{n,3} + T_{n,4} \quad \text{say.}
\end{aligned}$$

To complete the proof of the theorem, by Minkowski's inequality for  $k > 1$ , it is enough to show that

$$\sum_{n=1}^{\infty} (P_n/p_n)^{k+k-1} |T_{n,i}|^k < \infty, \quad \text{for } i = 1, 2, 3, 4. \quad (5.3)$$

Firstly, we have that

$$\begin{aligned}
\sum_{n=1}^m (P_n/p_n)^{k+k-1} |T_{n,1}|^k &= O(1) \sum_{n=1}^m |\lambda_n| (P_n/p_n)^{k-1} |t_n|^k \\
&= O(1) \sum_{n=1}^{m-1} \Delta |\lambda_n| \sum_{i=1}^n (P_i/p_i)^{k-1} |t_i|^k + O(1) |\lambda_m| \sum_{n=1}^m (P_n/p_n)^{k-1} |t_n|^k \\
&= O(1) \sum_{n=1}^{m-1} |\Delta \lambda_n| X_n + O(1) |\lambda_m| X_m = O(1) \sum_{n=1}^{m-1} A_n X_n + O(1) |\lambda_m| X_m = O(1)
\end{aligned}$$

as  $m \rightarrow \infty$ , by the hypotheses of the theorem and Lemma 1.

Now, applying Hölder's inequality with indices  $k$  and  $k'$ , where  $1/k + 1/k' = 1$ , as in  $T_{n,1}$  we have that

$$\begin{aligned}
\sum_{n=2}^{m+1} (P_n/p_n)^{k+k-1} |T_{n,2}|^k &= O(1) \sum_{n=2}^{m+1} (P_n/p_n)^{k-1} \frac{1}{P_{n-1}} \left\{ \sum_{v=1}^{n-1} P_v |t_v|^k |\lambda_v|^k \right\} \\
&\quad \times \left\{ \frac{1}{P_{n-1}} \sum_{v=1}^{n-1} P_v \right\}^{k-1} \\
&= O(1) \sum_{v=1}^m |\lambda_v|^{k-1} |\lambda_v| P_v |t_v|^k \sum_{n=v+1}^{m+1} (P_n/p_n)^{k-1} \frac{1}{P_{n-1}} \\
&= O(1) \sum_{v=1}^m |\lambda_v| (P_v/p_v)^{k-1} |t_v|^k = O(1) \quad \text{as } m \rightarrow \infty.
\end{aligned}$$

Using the fact that  $P_v = O(vp_v)$ , by (2.1), we have

$$\begin{aligned}
\sum_{n=2}^{m+1} (P_n/p_n)^{k+k-1} |T_{n,3}|^k &= \sum_{n=2}^{m+1} (P_n/p_n)^{k-1} \frac{1}{P_{n-1}^k} \left\{ \sum_{v=1}^{n-1} P_v |\Delta \lambda_v| |t_v| \right\}^k \\
&= O(1) \sum_{n=2}^{m+1} (P_n/p_n)^{k-1} \frac{1}{P_{n-1}^k} \left\{ \sum_{v=1}^{n-1} v p_v |A_v| |t_v| \right\}^k \\
&= O(1) \sum_{n=2}^{m+1} (P_n/p_n)^{k-1} \frac{1}{P_{n-1}} \left\{ \sum_{v=1}^{n-1} (v |A_v|)^k p_v |t_v|^k \right\} \left\{ \frac{1}{P_{n-1}} \sum_{v=1}^{n-1} P_v \right\}^{k-1}
\end{aligned}$$



$$\begin{aligned}
&= O(1) \sum_{v=1}^m (v|A_v|)^{k-1} v|A_v|p_v|t_v|^k \sum_{n=v+1}^{m+1} (P_n/p_n)^{y_{k-1}} \frac{1}{P_{n-1}} \\
&= O(1) \sum_{v=1}^m v|A_v|(P_v/p_v)^{y_{k-1}} |t_v|^k = O(1) \sum_{v=1}^{m-1} \Delta(v|A_v|) \sum_{i=1}^v (P_i/p_i)^{y_{k-1}} |t_i|^k \\
&\quad + O(1)m|A_m| \sum_{v=1}^m (P_v/p_v)^{y_{k-1}} |t_v|^k = O(1) \sum_{v=1}^{m-1} |\Delta(v|A_v|)|X_v \\
&\quad + O(1)m|A_m|X_m \\
&= O(1) \sum_{v=1}^{m-1} vX_v|\Delta A_v| + O(1) \sum_{v=1}^{m-1} |A_{v+1}|X_v + O(1)m|A_m|X_m = O(1)
\end{aligned}$$

as  $m \rightarrow \infty$ ,

by the hypotheses of the theorem and Lemma 2.

Finally, using the fact that  $P_v = O(vp_v)$ , by (2.1), as in  $T_{n,1}$  we have that

$$\begin{aligned}
&\sum_{n=2}^{m+1} (P_n/p_n)^{y_{k+k-1}} |T_{n,4}|^k \leq \sum_{n=2}^{m+1} (P_n/p_n)^{y_{k-1}} \frac{1}{P_{n-1}^k} \left\{ \sum_{v=1}^{n-1} \frac{P_v}{v} |\lambda_{v+1}| |t_v| \right\}^k \\
&= O(1) \sum_{n=2}^{m+1} (P_n/p_n)^{y_{k-1}} \frac{1}{P_{n-1}^k} \left\{ \sum_{v=1}^{n-1} p_v |\lambda_{v+1}| |t_v| \right\}^k \\
&= O(1) \sum_{n=2}^{m+1} (P_n/p_n)^{y_{k-1}} \frac{1}{P_{n-1}} \left\{ \sum_{v=1}^{n-1} p_v |\lambda_{v+1}|^k |t_v|^k \right\} \left\{ \frac{1}{P_{n-1}} \sum_{v=1}^{n-1} p_v \right\}^{k-1} \\
&= O(1) \sum_{v=1}^m |\lambda_{v+1}|^{k-1} |\lambda_{v+1}| p_v |t_v|^k \sum_{n=v+1}^{m+1} (P_n/p_n)^{y_{k-1}} \frac{1}{P_{n-1}} \\
&= O(1) \sum_{v=1}^m |\lambda_{v+1}| (P_v/p_v)^{y_{k-1}} |t_v|^k = O(1) \quad \text{as } m \rightarrow \infty.
\end{aligned}$$

Therefore, we obtain that

$$\sum_{n=1}^m (P_n/p_n)^{y_{k+k-1}} |T_{n,i}| = O(1) \quad \text{as } m \rightarrow \infty, \quad \text{for } i = 1, 2, 3, 4.$$

This completes the proof of the theorem. If we take  $p_n = 1$  for all values of  $n$  (in the case  $X_n \sim \log n$ ) in our theorem, then we get a result for  $|C, 1; \gamma|_k$  summability factor provided that  $1 - \gamma_k > 0$ . For  $\gamma = 0$ , this result due to Mazhar ([8]).

### Acknowledgement

This research was supported by TBAG-CG2 (Tübitak).

### References

- [2] Bor H, On two summability methods, *Math. Proc. Camb. Philos. Soc.* **97** (1985) 147–149
- [3] Bor H, A relation between two summability methods, *Riv. Mat. Univ. Parma* (4) **14** (1988) 107–112
- [4] Bor H, On quasi-monotone sequences and their applications, *Bull. Aust. Math. Soc.* **43** (1991) 187–192
- [5] Flett T M, Some more theorems concerning the absolute summability of Fourier series, *Proc. Lond. Math. Soc.* **8** (1958) 357–387
- [6] Hardy G H, *Divergent Series* (Oxford: University Press) (1949)
- [7] Kogbetliantz E, Sur les séries absolument sommables par la méthode des moyennes arithmétiques, *Bull. Sci. Math.* **49** (1925) 234–256
- [8] Mazhar S M, On generalized quasi-convex sequence and its applications, *Indian J. Pure Appl. Math.* **8** (1977) 784–790

## Rearrangements of bounded variation sequences

MEHMET ALI SARIGÖL

Department of Mathematics, Erciyes University, Kayseri 38039, Turkey

MS received 1 March 1993; revised 19 November 1993

**Abstract.** Let  $bv$  be the set of all bounded variation sequences. In the present paper we deduce from a theorem of Mears a necessary and sufficient condition for the rearrangement  $(a_{p(k)})$  to be of bounded variation whenever  $(a_k) \in bv$ ; interestingly it coincides with Pleasants' criterion for convergence-preserving.

**Keywords.** Rearrangements; sequences.

Let  $\Sigma a_k$  be an infinite series of real numbers and  $p$  be a permutation of  $N$ , the set of all positive integers. The series  $\Sigma a_{p(k)}$  is then called a rearrangement of  $\Sigma a_k$ . A classical theorem of Riemann states that if  $\Sigma a_k$  is a conditionally convergent series and  $s$  is any fixed real number (or  $\pm \infty$ ), then there is a permutation  $p$  such that  $\Sigma a_{p(k)} = s$ . Thus it leads us to the problem of characterizing the rearrangements which do not change the sum or the convergence or even the divergence of the series. They were studied in [1]–[9] and by others. Of special interest is a paper by Pleasants [5] giving a characterization of permutations which transform convergent sequences to convergent sequences.

In this paper we consider questions similar to those above, but for rearrangements of bounded variation sequences. We recall some notation before stating the precise problem.

### DEFINITION

Let  $\gamma$  denote the set of all convergent series of real numbers. A permutation  $p$  on positive integers is then called convergence-preserving (CP for short) if  $a_p = (a_{p(k)}) \in \gamma$  for any sequence  $a = (a_k) \in \gamma$ , [5].

We shall denote the finite consecutive run of integers  $i, i+1, \dots, j-1, j$  by  $[i, j]$  and we shall call such a set a block.

Every finite set  $F$  of  $N$  is a union of disjoint, non-adjacent blocks of consecutive integers. Let  $v(F)$  denote the number of such blocks and  $p^{-1}$  be the inverse of  $p$ . With this terminology, the result of Pleasants on CP permutations can be stated as follows.

**Theorem 1.** [5; p. 135]. *A permutation  $p$  is CP if and only if there is a constant  $M = M(p)$  such that  $v(p^{-1}\{1, 2, \dots, k\}) \leq M$  for all  $k \in N$ .*

We now give the following notation similar to the above definition, for sequences of bounded variation.

## DEFINITION

Let  $p$  be a permutation of  $N$ . We say that  $p$  is bounded variation-preserving (written BVP) if the rearrangement induced by  $p$ ,  $a_p = (a_{p(k)})$ , is of bounded variation for all  $a = (a_k) \in bv$ .

It may be noticed that, if  $a \in bv$  then any rearrangement of  $a$  is not necessarily of bounded variation. For example, take  $a = (1, 1/2, \dots, 1/n, \dots) \in bv$  and define permutation  $p$  as

$$p(n) = \begin{cases} 2n/3, & \text{if } n \equiv 0 \pmod{3} \\ (4n-1)/3, & \text{if } n \equiv 1 \pmod{3} \\ (4n+1)/3, & \text{if } n \equiv 2 \pmod{3}. \end{cases}$$

Then  $(a_{p(n)}) \in bv$ , since

$$\sum_{n=2}^{\infty} |a_{p(n)} - a_{p(n-1)}| > \sum_{n \equiv 1 \pmod{3}} 3(2n+1)/2(n-1)(4n-1) = \infty.$$

In this paper we give the following characterization of such permutations and obtain a relation between CP and BVP permutations.

**Theorem 2.** *A permutation  $p$  is BVP if and only if there exists a positive integer  $M$  such that  $v(p^{-1}\{1, 2, \dots, k\}) \leq M$  for all  $k$ .*

We deduce this from the following result of Mears [5] on bounded variation sequences.

**Theorem 3.** *The infinite matrix  $A = (a_{nk})$  transforms every  $x \in bv$  into  $((Ax)_n) \in bv$ , where  $(Ax)_n = \sum a_{nk} x_k$  if and only if*

- (i)  $\sum_{k=1}^{\infty} a_{nk}$  converges for each  $n$ , and
- (ii) There exists a positive constant  $M$  such that, for all  $k$ ,

$$\sum_{n=1}^{\infty} \left| \sum_{i=k}^{\infty} (a_{ni} - a_{n-1,i}) \right| \leq M.$$

## Proof of Theorem 2

Suppose that  $p$  is any permutation on  $N$  and  $x \in bv$ . A rearrangement of  $x$  by  $p$  can be considered as a matrix transformation in the following way. Set  $a_{nk} = 1$  if  $k = p(n)$ , and  $a_{nk} = 0$  if  $k \neq p(n)$ . Then  $(Ax)_n = x_{p(n)}$ , i.e.,  $Ax = (x_{p(n)})$ , the rearrangement produced by  $p$ . Since  $p$  is one-to-one and onto mapping, each column and each row of  $(a_{nk})$  contains exactly one nonzero term. Thus for each  $n$ ,  $\sum_{k=1}^{\infty} a_{nk} = 1$  and also for each  $k > 1$ ,

$$\sum_{n=1}^{\infty} \left| \sum_{i=k}^{\infty} (a_{ni} - a_{n-1,i}) \right| = \sum_{n=1}^{\infty} \left| \sum_{i=1}^{k-1} (a_{ni} - a_{n-1,i}) \right|.$$

Since  $p^{-1}$  is also the permutation on  $N$ ,  $1 \in \{p^{-1}(1), p^{-1}(2), \dots, p^{-1}(k)\}$  for all sufficiently large  $k$ . Now arranging the integers in  $\{p^{-1}(1), \dots, p^{-1}(k)\}$  in increasing order, we

write the same set in the following form:

$$\{1 = r_0^{(k)}, 2, \dots, t_0^{(k)}, r_1^{(k)}, r_1^{(k)} + 1, \dots, t_1^{(k)}, r_2^{(k)}, r_2^{(k)} + 1, \dots, t_2^{(k)}, \dots, r_{j_k}^{(k)}, r_{j_k}^{(k)} + 1, \dots, t_{j_k}^{(k)}\}, \quad (1)$$

where

$$r_0^{(k)} \leq t_0^{(k)} < r_1^{(k)} - 1 < t_1^{(k)} < r_2^{(k)} - 1 < \dots < r_{j_k}^{(k)} - 1 < t_{j_k}^{(k)}.$$

It is clear that  $v(p^{-1}[1, k]) = j_k + 1$ . Because of the definition of  $p$ , there is only one nonzero term in each row corresponding to the integers in (1), and so

$$\sum_{i=1}^k a_{ni} = \begin{cases} 1, & \text{if } r_j^{(k)} \leq n \leq t_j^{(k)}, \quad j = 0, 1, \dots, j_k \\ 0, & \text{otherwise,} \end{cases}$$

which implies that

$$\sum_{i=1}^k (a_{ni} - a_{n-1,i}) = \begin{cases} -1, & \text{if } n = t_j^{(k)} \\ 1, & \text{if } n = r_j^{(k)}, \quad j = 0, 1, \dots, j_k \\ 0, & \text{otherwise} \end{cases}$$

Therefore it follows that

$$\sum_{n=1}^{\infty} \left| \sum_{i=1}^k (a_{ni} - a_{n-1,i}) \right| = 2(j_k + 1),$$

because the above sum is '1' added to itself  $2(j_k + 1)$  times. So, by Theorem 3,  $Ax = ((Ax)_n) \in bv$  iff there is a constant  $M$  such that  $j_k \leq M$  for all  $k$ . This completes the proof.

#### COROLLARY 4

*A permutation is CP if and only if it is BVP.*

In [5] it is shown that there exists permutations  $p$  which are CP but the inverse  $p^{-1}$  not CP. In particular it follows that.

#### COROLLARY 5

*The set of all BVP permutations does not form a group.*

#### Acknowledgement

The author expresses his gratitude to the referee for valuable suggestions. This research was supported by TBAG-ÇG2 (TÜBİTAK).

#### References

- [3] Levi F W, Rearrangement of convergent series, *Duke Math. J.* **13** (1946) 579–585
- [4] Mears F M, Absolutely regularity and Nörlund mean, *Ann. Math.* **38** (1938) 594–601
- [5] Pleasants P A B, Rearrangements that preserve convergence, *J. London Math. Soc.* **15** (1977) 134–142
- [6] Sarigöl M A, Permutation preserving convergence and divergence of series, *Bull. Inst. Math. Acad. Sinica* **16** (1988) 221–227
- [7] Sarigöl M A, On absolute equivalence of permutation functions, *Bull. Inst. Math. Acad. Sinica* **19** (1991) 69–74
- [8] Sarigöl M A, A short proof of Levi's theorem on rearrangements of convergent series, *Doğa-Tr. J. Math.* **16** (1992) 201–205
- [9] Schaeier P, Sum-preserving rearrangements of infinite series, *Am. Math. Mon.* **88** (1981) 33–40

# A note on a generalization of Macdonald's identities for $A_\ell$ and $B_\ell$

N STHANUMOORTHY and M TAMBA\*

The Ramanujan Institute, University of Madras, Madras 600 005, India

\*School of Mathematics, SPIC Science Foundation, Madras 600 017, India

MS received 18 June 1990; revised 4 September 1993

**Abstract.** Let  $\eta(q)$  denote the Dedekind's  $\eta$ -function. Macdonald obtained identities for  $\eta(q)^{\dim g}$  where  $g$  is complex simple finite dimensional Lie algebra. The aim of this paper is to obtain generalization of the above identities in the case of  $g = A_\ell$  and  $B_\ell$ . We also get new formulas for the generating functions of the Ramanujan's  $\tau$ -function and  $\psi_\alpha$ -functions.

**Keywords.** Macdonald's multivariable identities; Dedekind's eta function;

## 1. Introduction

Let  $\eta(q)$  denote Dedekind's  $\eta$ -function. Macdonald [7] obtained a formula for  $\eta(q)^{\dim g}$  for every complex simple Lie algebra  $g$  which gives a generalization of the Jacobi's expansion for  $\eta(q)^3$ . These formulas are some specializations of the Macdonald's multivariable identities [7]. Many other identities involving Dedekind's  $\eta$ -function were also obtained by Lepowsky [4].

We shall need the following preliminaries. Let  $g$  denote the simple finite dimensional Lie algebra of the type  $A_\ell$  or  $B_\ell$ ;  $\mathfrak{h}$  denotes its Cartan subalgebra; Let  $\Delta$ ,  $\Delta_+$  and  $\Delta_+^\vee$  be the root system, the set of positive roots and the positive dual roots respectively of  $g$  and  $\rho$  (respectively  $\rho^\vee$ ) be the half-sum of the roots in  $\Delta_+$  (resp.  $\Delta_+^\vee$ ).  $W$  and  $M$  are the Weyl group of  $g$  and the lattice spanned over  $\mathbb{Z}$  by the long roots of  $g$  respectively. Let  $h$  and  $g$  denote the Coxeter and dual Coxeter numbers of  $g$  respectively.

Let  $\langle, \rangle$  denote the pairing of the elements in  $\mathfrak{h}$  and  $\mathfrak{h}^*$  (cf. [3]). We introduce the following notations:

$$\Delta_m^\vee = \{\alpha \in \Delta_+^\vee / \langle \rho, \alpha \rangle \equiv 0 \pmod{m}\},$$
$$d_m(\lambda) = \prod_{\alpha \in \Delta_m^\vee} \frac{\langle \lambda + \rho, \alpha \rangle}{\langle \rho, \alpha \rangle}, \quad (\lambda \in \mathfrak{h}^*).$$

The aim of this paper is to prove the following:

**Theorem 1.** For  $\ell \geq 1$ , let  $m \leq \ell + 1$  be any divisor of  $\ell + 1$ . Then

$$\eta(q^m)^{(\ell+1)^2/m} \eta(q)^{-1} = \sum_{\alpha \in M} d_m((\ell+1)\alpha)^{(1/2(\ell+1))|\rho + (\ell+1)\alpha|^2}$$

where

$$d_m(\lambda) = \prod_{\alpha \in \Delta_m^\vee} \frac{\langle \lambda + \rho, \alpha \rangle}{\langle \rho, \alpha \rangle}, (\lambda \in \mathfrak{h}^*).$$

**Theorem 2.** For  $\ell \geq 3$ , let  $m$  be any divisor of  $(2\ell - 1)$ . Then

$$\eta(q)\eta(q^m)^{(2\ell^2 + \ell - 1)/m} = \sum_{\alpha \in M} d_m((2\ell - 1)\alpha) q^{(1/2(2\ell - 1)|\rho + (2\ell - 1)\alpha|^2)}.$$

The above identities are generalizations of Macdonald's formula for  $\eta(q)^{(\ell+1)^2-1}$  and  $\eta(q)^{2\ell^2+\ell}$ .

The consequence of the above theorems are the following corollaries.

**COROLLARY 1**

$$\eta(q)^{24} = \sum_{\substack{r_1, \dots, r_{12}; r \in \mathbb{Z} \\ r_1 + \dots + r_{12} = 0}} \left\{ (-1)^r \cdot D_1(r_1, \dots, r_{12}) \cdot q^{(1/44)((6r+1)^2)} + \sum_{i=1}^{12} (13 - 2i + 24r_i)^2/4 \right\}$$

where

$$D_1(r_1, \dots, r_{12}) = \prod_{i=1}^6 (2(r_i - r_{i+6}) + 1).$$

**COROLLARY 2**

$$\begin{aligned} & q\{(1-q^3)(1-q^6)\dots\}^8 \\ &= \sum_{\substack{r_1, \dots, r_4; r \in \mathbb{Z} \\ r_1 + \dots + r_4 = 0}} \left\{ (-1)^r \cdot D_2(r_1, \dots, r_4) \cdot q^{(1/16)((6r+1)^2 + (3/4))} \sum_{i=1}^4 (5 - 2i + 8r_i)^2 \right\}, \end{aligned}$$

where

$$D_2(r_1, \dots, r_4) = (2(r_1 - r_3) + 1)(2(r_2 - r_4) + 1).$$

**COROLLARY 3**

$$\begin{aligned} & q\{(1-q)^2 \cdot (1-q^4)\dots\}^{12} \\ &= \sum_{\substack{r_1, \dots, r_6; r \in \mathbb{Z} \\ r_1 + \dots + r_6 = 0}} \left\{ (-1)^r \cdot D_3(r_1, \dots, r_6) \cdot q^{(1/36)((6r+1)^2 + (1/2))} \sum_{i=1}^6 (7 - 2i + 12r_i)^2 \right\}, \end{aligned}$$

where

$$D_3(r_1, \dots, r_6) = (2(r_1 - r_4) + 1)(2(r_2 - r_5) + 1)(2(r_3 - r_6) + 1).$$

**COROLLARY 4**

For  $\ell \geq 3$ , we have

$$\eta(q)\eta(q^{2\ell-1})^{\ell+1} = \sum_{\alpha \in M} d_{2\ell-1}((2\ell-1)\alpha) q^{(1/2(2\ell-1)|\rho + (2\ell-1)\alpha|^2)}.$$



Corollary 1 gives a new formula for the generating function  $\eta(q)^{24}$  of the Ramanujan's  $\tau$ -function. This is different from that of Dyson [1] and Lepowsky [4, 5]. Corollaries 2 and 3 give formulas for the generating function  $G_k$  of the Ramanujan's  $\psi_\alpha$ -functions [8].

Now we briefly explain the techniques used to obtain the above results: The substitutions involved in these computations are actually generalizations of the substitutions used by Macdonald [7] and Lepowsky [4, 5].

## 2. Explanation of the techniques involved

We will use the following form of the Macdonald's identity for the affine Lie algebra of the type  $A_\ell^{(1)}$  or  $B_\ell^{(1)}$  (the identity is true for any  $X_\ell^{(1)}$ ) (cf. [3]: pp. 168):

$$\begin{aligned} e\left(-\frac{|\rho|^2}{2g}\delta\right) \prod_{n \geq 1} ((1 - e(-n\delta))^\ell \cdot \prod_{\alpha \in \Delta_-} (1 - e(-n\delta + \alpha))) \\ = \sum_{\alpha \in M} \chi(g\alpha) e\left(-\frac{1}{2g}|\rho + g\alpha|^2\delta\right), \end{aligned} \quad (2.1)$$

where, for  $\lambda \in \mathfrak{h}^*$

$$\chi(\lambda) = \frac{\sum_{w \in W} \varepsilon(w) e(w(\lambda + \rho) - \rho)}{\prod_{\alpha \in \Delta_+} (1 - e(-\alpha))}. \quad (2.2)$$

Let  $\alpha_1, \dots, \alpha_\ell$  denote the simple roots of  $\mathfrak{g}$ . For any divisor  $m$  of  $g$ , where  $g$  stands for the dual Coxeter number of  $\mathfrak{g}$ , let  $\phi_m$  denote the specialization  $\phi_m(e(-\delta)) = q$ ,  $\phi_m(e(-\alpha_i)) = w(i = 1, \dots, \ell)$ ; here  $w$  denotes the primitive root of degree  $m$  of unity.

We require the following Lemma.

*Lemma.* For  $\alpha \in M$ ,

$$\phi_m(\chi(g\alpha)) = d_m(g\alpha). \quad (2.3)$$

*Proof.* Fix  $\alpha \in M$ . We define the following homomorphisms:

$$F_1: \mathbb{C}[[e(-\alpha_i): 1 \leq i \leq \ell]] \rightarrow \mathbb{C}[[t, t^{-1}]]$$

by

$$F_1(e(-\alpha_i)) = t \quad (i = 1, 2, \dots, \ell).$$

(Here note that  $F_1(e(-\alpha)) = t^{\langle \alpha, \rho^\vee \rangle}$ ) and

$$F_2: \mathbb{C}[[e(-\alpha_i^\vee): 1 \leq i \leq \ell]] \rightarrow \mathbb{C}[[t, t^{-1}]]$$

by

$$F_2(e(-\alpha_i^\vee)) = t^{\langle \alpha_i^\vee, g\alpha + \rho \rangle}.$$

Since  $\phi_m$  is an homomorphism and since  $\lim_{t \rightarrow w} F_1(e(-g\alpha)) = 1$ , it suffices to prove that

$$\phi_m(e(-g\alpha)) \chi(g\alpha) = d_m(g\alpha).$$

Now,

$$\begin{aligned}
 \phi_m(e(-g\alpha)\chi(g\alpha)) &= \phi_m \left\{ \frac{\sum_{w \in W} \varepsilon(w) e(w(g\alpha + \rho) - (\rho + g\alpha))}{\prod_{\beta \in \Delta_+} (1 - e(-\beta))} \right\} \\
 &= \lim_{t \rightarrow \omega} \left\{ \frac{\sum_{w \in W} \varepsilon(w) t^{\langle g\alpha + \rho, \rho^\vee \rangle - \langle w(g\alpha + \rho), \rho^\vee \rangle}}{\prod_{\beta \in \Delta_+} (1 - t^{\langle \beta, \rho^\vee \rangle})} \right\} \\
 &= \lim_{t \rightarrow \omega} \left\{ \frac{F_2(\sum_{w \in W} \varepsilon(w) e(w(\rho^\vee) - \rho^\vee))}{F_1(\prod_{\beta \in \Delta_+} (1 - e(-\beta)))} \right\} \\
 &= \lim_{t \rightarrow \omega} \frac{F_2(\prod_{\beta \in \Delta_+^\vee} (1 - e(-\beta)))}{F_1(\prod_{\beta \in \Delta_+} (1 - e(-\beta)))}, \quad (\text{by [3], 10.4.4}) \\
 &= \lim_{t \rightarrow \omega} \frac{(\prod_{\beta \in \Delta_+^\vee} (1 - t^{\langle g\alpha + \rho, \beta \rangle}))}{\prod_{\beta \in \Delta_+} (1 - t^{\langle \beta, \rho^\vee \rangle})} \\
 &= \lim_{t \rightarrow \omega} \left\{ \frac{\prod_{\beta \in \Delta_+^\vee} (1 - t^{\langle g\alpha + \rho, \beta \rangle})}{\prod_{\beta \in \Delta_+} (1 - t^{\langle \beta, \rho^\vee \rangle})} \right\} \\
 &= d_m(g\alpha) \quad (\text{by L' Hospital's rule})
 \end{aligned}$$

Let  $[x]$  denote the greatest integer contained in  $x$ . Let  $\eta_p$  denote the number of roots in  $\Delta_+$  with height  $p$  [5] and for  $0 \leq j \leq m$ , let  $N_j(m)$  denote the number of roots in  $\Delta$  with height congruent to  $j \pmod{m}$ . It is not hard to see that

$$N_0(m) = 2 \sum_{k=1}^{[h/m]} \eta_{km} \quad (\text{for } j=0) \quad (2.4)$$

and

$$N_j(m) = \eta_j + \sum_{k=1}^{[h/m]-1} (\eta_{km-j} + \eta_{km+j}) + \eta_{[h/m]m-j}, \quad (\text{for } 1 \leq j < m). \quad (2.5)$$

Now applying  $\phi_m$  to (2.1) and using (2.3), (2.4) and (2.5) along with the strange formula of Freudenthal de varies:

$$\frac{|\rho|^2}{2g} = \frac{\dim \mathfrak{g}}{24} \quad (\text{cf. [3]}), \quad (2.6)$$

we obtain

$$q^{\dim \mathfrak{g}/24} \prod_{n \geq 1} \left( (1 - q^n)^{\ell} \prod_{j=0}^{m-1} (1 - q^n \omega^j)^{N_j(m)} \right) = \sum_{\alpha \in M} d_m(g\alpha) q^{(1/2g)[\rho + g\alpha]^2}. \quad (2.7)$$

Now, using the known facts about  $\eta_p$  (cf. [5; pp. 228]) one can easily compute  $\eta_p$ :

$$\eta_p = \ell + 1 - p \quad \text{for } g = A_\ell$$

and

Case (i).  $g$  is of type  $A_\ell$ . By (2.4) and (2.5) we have:

$$N_0(m) = ((\ell + 1)^2/m) - (\ell + 1), \quad (2.8)$$

and for  $1 \leq j < m$ ,

$$N_j(m) = (\ell + 1)^2/m. \quad (2.9)$$

Case (ii).  $g$  is of type  $B_\ell$  and  $m = 1$ .

In this case  $\left[\frac{h}{m}\right] = 2\ell$ . Hence by (2.4) we have

$$\begin{aligned} N_0(1) &= 2 \left\{ \sum_{\substack{k=1 \\ (k \text{ odd})}}^{2\ell} \eta_k + \sum_{\substack{k=1 \\ (k \text{ odd})}}^{2\ell} \eta_k \right\} \\ &= 2 \left\{ \sum_{\substack{k=1 \\ (k \text{ odd})}}^{2\ell} (\ell - (k-1)/2) + \sum_{\substack{k=1 \\ (k \text{ even})}}^{2\ell} (\ell - k/2) \right\} \\ &= (2\ell)(2\ell) - (2\ell)(2\ell + 1)/2 + \ell \\ &= 2\ell^2 \end{aligned} \quad (2.10)$$

Case (iii).  $g$  is of type  $B_\ell$  and  $m > 1$ . In this case  $\left[\frac{h}{m}\right] = \frac{(2\ell - 1)}{m}$ . Hence by (2.4) we have

$$\begin{aligned} N_0(m) &= 2 \left\{ \sum_{\substack{k=1 \\ (k \text{ odd})}}^{(2\ell-1)/m} \eta_{km} + \sum_{\substack{k=1 \\ (k \text{ even})}}^{(2\ell-1)/m} \eta_{km} \right\}, \\ &= 2 \left\{ \sum_{\substack{k=1 \\ (k \text{ odd})}}^{(2\ell-1)/m} (\ell - (km-1)/2) + \sum_{\substack{k=1 \\ (k \text{ even})}}^{(2\ell-1)/m} (\ell - (km)/2) \right\}, \\ &= ((2\ell - 1)/m)(2\ell) - m((2\ell - 1)/m)((2\ell - 1)/m + 1)/2 + ((2\ell - 1)/m + 1)/2 \\ &= (2\ell^2 + \ell - 1)/m - (\ell - 1). \end{aligned} \quad (2.11)$$

Furthermore, one can easily see that

$$\eta_{km-j} + \eta_{km+j} = \begin{cases} 2\ell - km & \text{if } j \text{ is even and } k \text{ is even,} \\ & \text{or } j \text{ is odd and } k \text{ is odd.} \\ 2\ell - km + 1 & \text{if } j \text{ is even and } k \text{ is odd,} \\ & \text{or } j \text{ is odd and } k \text{ is even.} \end{cases}$$

Hence we have by (2.5), that for  $1 \leq j < m$  and  $j$  even,

$$N_j(m) = \eta_j + \left\{ \sum_{\substack{k=1 \\ (k \text{ even})}}^{((2\ell-1)/m)-1} (\eta_{km-j} + \eta_{km+j}) + \sum_{\substack{k=1 \\ (k \text{ odd})}}^{((2\ell-1)/m)-1} (\eta_{km-j} + \eta_{km+j}) \right\}$$

$$+ \eta_{((h/m)m-j)}$$

$$\begin{aligned}
&= (\ell - j/2) + \sum_{\substack{k=1 \\ (k \text{ even})}}^{((2\ell-1)/m)-1} (2\ell - km) + \sum_{\substack{k=1 \\ (k \text{ odd})}}^{((2\ell-1)/m)-1} (2\ell - km + 1) \\
&\quad + (\ell - ((2\ell-1) - j - 1)/2) \\
&= (2\ell)(2\ell-1)/m - \frac{1}{2}((2\ell-1)/m-1)((2\ell-1)/m)m + ((2\ell-1)/m-1)2 - (\ell-1), \\
&= (2\ell^2 + \ell - 1)m.
\end{aligned} \tag{2.12}$$

Similarly, for  $1 \leq j < m$  and  $j$  odd:

$$\begin{aligned}
N_j(m) &= N_j + \sum_{\substack{k=1 \\ (k \text{ even})}}^{((2\ell-1)/m)-1} (\eta_{km-j} + \eta_{km+j}) + \sum_{\substack{k=1 \\ (k \text{ odd})}}^{((2\ell-1)/m)-1} (\eta_{km-j} + \eta_{km+j}) + \eta_{(2\ell-1)-j} \\
&= (\ell - (j-1)/2) + \sum_{\substack{k=1 \\ (k \text{ even})}}^{((2\ell-1)/m)-1} (2\ell - km + 1) + \sum_{\substack{k=1 \\ (k \text{ odd})}}^{((2\ell-1)/m)-1} (2\ell - km) \\
&\quad + (\ell - ((2\ell-1) - j)/2) \\
&= (2\ell^2 + \ell - 1)/m.
\end{aligned} \tag{2.13}$$

Now, using (2.4), (2.5) along with (2.9), (2.10), (2.12), (2.13) and the fact that

$$\prod_{j=1}^{n-1} (1 - a\omega^j) = (1 - a^n)(1 - a)^{-1}, \quad (a \neq 1)$$

theorems 1 and 2 follow.

Note that for  $m = 1$  the identities of theorems 1 and 2 are precisely the Macdonald's identities for  $\eta(q)^{(\ell+1)^2-1}$  and  $\eta(q)^{2\ell^2+\ell}$  respectively.

Furthermore, using the following identity due to Euler (cf. [6]):

$$\eta(q) = \sum_{r \in \mathbb{Z}} (-1)^r q^{(1/24)(6r+1)^2}$$

and by replacing  $q$  by  $q^{24/(\ell+1)^2}$  in theorem 1, we obtain:

$$\eta(q^{24m/(\ell+1)^2})^{(\ell+1)^2/m} = \sum_{\substack{\alpha \in M \\ r \in \mathbb{Z}}} (-1)^r d_m((\ell+1)\alpha) q^{(12/(\ell+1)^3)\{(\ell\alpha + (\ell+1)\alpha^2) + 1/(\ell+1)^2(6r+1)^2\}}. \tag{2.14}$$

Now the Corollaries 1, 2 and 3 follow by taking  $n = 12$  and  $m = 6$ ,  $n = 4$  and  $m = 2$ , and  $n = 6$  and  $m = 3$  respectively. Corollary 4 follows by taking  $m = 2l - 1$ , by Theorem 2.

### Acknowledgement

The authors are very much thankful to the referee for valuable suggestions and helpful comments. The second author (MT) wishes to thank the National Board for Higher Mathematics for the financial support.

**References**

- [1] Dyson F J, Missed opportunities, *Bull. Math. Soc.* **78** (1972) 635–652
- [2] Hardy G H, *Ramanujan* Chelsea Publishing Company, New York (1940)
- [3] Kac V G, *Infinite-dimensional Lie algebras* (2nd ed.), (Cambridge, University Press) (1985)
- [4] Lepowsky J, Macdonald-type identities, *Adv. Math.* **27** (1978) 230–234
- [5] Lepowsky J, Generalized Verma modules, Loop space cohomology and Macdonald-type identities, *Ann. Sci. Ecole. Norm. Sup.* **12** (1979) 169–234
- [6] Lepowsky J, Affine Lie algebras and combinatorial identities, Lie algebras and Related Topics, *Lecture Notes in Mathematics* (Springer-Verlag) 933 (1982) 130–156.
- [7] Macdonald I G, Affine root systems and Dedekind's  $\eta$ -function, *Invent. Math.* **15** (1972) 91–143
- [8] Ramanujan S, On certain arithmetical functions, *Trans. Cambridge Philos. Soc.* **22** (1916) 159–184



## Combinatorial manifolds with complementarity

BASUDEB DATTA

Department of Mathematics, Indian Institute of Science, Bangalore 560 012, India

MS received 2 July 1993; revised 25 September 1993.

**Abstract.** A simplicial complex is said to satisfy complementarity if exactly one of each complementary pair of nonempty vertex-sets constitutes a face of the complex.

We show that if a  $d$ -dimensional combinatorial manifold  $M$  with  $n$  vertices satisfies complementarity then  $d = 0, 2, 4, 8$  or  $16$  with  $n = 3d/2 + 3$  and  $|M|$  is a “manifold like a projective plane”. Arnoux and Marin had earlier proved the converse statement.

**Keywords.** Combinatorial manifolds; complementarity.

### 1. Introduction

Recall that a *simplicial complex*  $K$  is a collection of nonempty sets (sets of *vertices*) such that all nonempty subsets of a member of the collection are again members. A member of  $K$  with  $i + 1$  vertices is called an  $i$ -*face* (or simplex of dimension  $i$ ). For  $\sigma \in K$   $\text{Lk}(\sigma) (= \text{link of } \sigma) := \{\gamma \in K; \gamma \cap \sigma = \emptyset, \gamma \cup \sigma \in K\}$ . A simplicial complex may be thought of as a prescription for the construction of a topological space by pasting together geometric simplexes. The topological space thus obtained from a simplicial complex  $K$  is called a *polyhedron* and is denoted by  $|K|$ . Let  $K_1$  and  $K_2$  be two simplicial complexes. A map  $f: |K_1| \rightarrow |K_2|$  is called PL if there are subdivisions  $K'_1$  and  $K'_2$  of  $K_1$  and  $K_2$  respectively such that  $f: K'_1 \rightarrow K'_2$  is simplicial. We write  $|K_1| \approx |K_2|$  if  $|K_1|$  and  $|K_2|$  are PL homeomorphic. A simplicial complex  $K$  (respectively  $|K|$ ) is called a *combinatorial  $d$ -manifold* (respectively *PL  $d$ -manifold*) if for every vertex  $v$  in  $K$   $\text{Lk}(v)$  is a  $(d - 1)$ -dimensional combinatorial sphere.

In 1962, Eells and Kuiper [5] proved that a PL manifold  $M^d$  with PL Morse number  $\mu(M^d) = 3$  has dimension  $d = 0, 2, 4, 8$  or  $16$ . If  $d = 0$   $M^d$  consists of three points. If  $d = 2$   $M^d$  is the real projective plane. For  $d = 4, 8$  or  $16$ ,  $M^d$  is a simply connected cohomology projective plane over complex numbers, quaternions or Cayley numbers, respectively. Each of the manifolds of above type is called a *manifold like a projective plane*. This classification turned up in the 1987 paper [3] of Brehm and Kühnel on combinatorial manifolds with few vertices. Specifically, they proved that: Let  $M_n^d$  be a combinatorial  $d$ -manifold with  $n$  vertices,

(BK1) if  $n < 3[d/2] + 3$  then  $|M_n^d| \approx S^d$ ,

(BK2) if  $n = 3(d/2) + 3$  and  $|M_n^d| \not\approx S^d$  then  $d = 2, 4, 8$  or  $16$  and  $|M_n^d|$  must be a “manifold like a projective plane”. Moreover for  $d = 2$   $M_n^d = \mathbb{R}P^2_6$  and for  $d = 4$   $M_n^d = \mathbb{C}P^2$

It is classically known that there exists a unique (up to simplicial isomorphism) 6-vertex triangulation (denoted by  $\mathbb{R}P_6^2$ ) of the real projective plane  $\mathbb{R}P^2$ . It is also known (see [2], [6] and [7]) that there exists a unique (up to simplicial isomorphism) 9-vertex triangulation (denoted by  $\mathbb{C}P_9^2$ ) of the complex projective plane  $\mathbb{C}P^2$ .

Implicit in [3] is the result that  $\mathbb{C}P_9^2$  satisfies complementarity. This result was made explicit by Arnoux and Marin [1] in 1991. More generally, they proved that any manifold as in (BK2) satisfies complementarity. In this article we prove the converse:

**Theorem.** Let  $M_n^d$  be a combinatorial  $d$ -manifold with  $n$  vertices. If  $M_n^d$  satisfies complementarity then  $d = 0, 2, 4, 8$  or  $16$  with  $n = 3(d/2) + 3$  and  $|M_n^d|$  is a "manifold like a projective plane".

## 2. Preliminaries

Let  $K$  be a triangulation of the sphere  $S^{p-1}$  with  $n$  vertices. The  $f$ -vector of  $K$  is  $f(K) := (f_0, \dots, f_{p-1})$ , where  $f_i$  is the number of  $i$ -faces in  $K$ . Thus  $f_0 = n$  and  $f_i \leq \binom{n}{i+1}$  for  $1 \leq i \leq p-1$ . Let  $\mathbb{N}$  denote the non-negative integers, and define  $H: \mathbb{N} \rightarrow \mathbb{N}$  as follows

$$H(m) = \begin{cases} 1 & \text{if } m = 0 \\ \sum_{i=0}^{p-1} f_i \binom{m-1}{i} & \text{if } m > 0. \end{cases} \quad (1)$$

Then there exists (see [8]) integers  $h_0, \dots, h_p$  such that

$$(1-x)^p \sum_{m=0}^{\infty} H(m)x^m = h_0 + h_1x + \dots + h_px^p \quad (2)$$

is an identity in the formal power series ring  $\mathbb{C}[[x]]$ .

For  $k \leq p < n-1$  (equating the coefficients of  $x^k$  from both sides of  $(1+x)^{-(p-k+1)}(1+x)^n = (1+x)^{n+k-p-1}$ ) we get

$$\sum_{j=0}^k (-1)^{k-j} \binom{p-j}{p-k} \binom{n}{j} = \binom{n+k-p-1}{k}. \quad (3)$$

By substituting  $i-1 = p$  and  $l = p+1-k$  we get

$$\binom{n-l}{i-l} = \sum_{j=0}^{i-l} (-1)^{i-l-j} \binom{i-1-j}{l-1} \binom{n}{j} = \sum_{m=l}^i (-1)^{i-m} \binom{n}{i-m} \binom{m-1}{l-1}. \quad (4)$$

Then from (1) and (2) by using (4) we get (see [9])

$$h_i = \sum_{l=0}^p (-1)^{i-l} \binom{p-l}{p-i} f_{l-1}, \quad (5)$$

where we set  $f_{-1} = 1$ .



If  $f_{j-1} = \binom{n}{j}$  for  $1 \leq j \leq q \leq p$  then by (3) we have

$$h_i = \binom{n+i-p-1}{i} \quad \text{for } i \leq q. \quad (6)$$

The Dehn-Sommerville equations, which hold for any triangulation of the sphere  $S^{p-1}$ , are equivalent to the statement (see [9]):

$$h_i = h_{p-i} \quad 0 \leq i \leq p. \quad (7)$$

### 3. Proof of the theorem

Throughout,  $M$  is an  $n$ -vertex combinatorial  $d$ -manifold satisfying complementarity. It is trivial from the definition that, for  $d=0$   $M$  consists of three points, and since clearly there is no 1-manifold satisfies complementarity, we may take  $d \geq 2$ .

We shall repeatedly use the following obvious consequences of complementarity. Since no set of  $\geq d+2$  vertices constitute a face,  $n \leq 2d+3$  and every set of  $\leq n-d-2$  vertices is a face. That is, for  $i \leq n-d-3$ , all  $i$ -faces occur in  $M$ . More generally the number of  $i$ -faces + the number of  $(n-i-2)$ -faces =  $\binom{n}{i+1}$ . As each vertex forms a 0-face, therefore  $n > d+2$ . Thus,  $d+2 < n \leq 2d+3$ .

Throughout this section we put  $c = [d/2]$ . Thus,  $d = 2c-1$  or  $2c$ .

If  $F_i$  is the number of  $i$ -faces in  $M$  then we have:

$$\begin{aligned} \sum_{i=0}^{n-3} F_i &= \begin{cases} F_0 + (F_1 + F_{2m-3}) + \cdots + (F_{m-2} + F_m) + F_{m-1} & \text{if } n = 2m, \\ F_0 + (F_1 + F_{2m-2}) + \cdots + (F_{m-1} + F_m) & \text{if } n = 2m+1 \end{cases} \\ &= \begin{cases} \binom{2m}{1} + \binom{2m}{2} + \cdots + \binom{2m}{m-1} + \frac{1}{2} \binom{2m}{m} & \text{if } n = 2m, \\ \binom{2m+1}{1} + \binom{2m+2}{2} + \cdots + \binom{2m+1}{m} & \text{if } n = 2m+1 \end{cases} \\ &= 2^{n-1} - 1, \end{aligned}$$

which is an odd integer, where we set  $F_i = 0$  for  $i > d$ . Therefore the Euler characteristic of  $M = \sum_{i=0}^{n-3} (-1)^i F_i$  is odd.

If  $n = d+3$  then (by (BK1))  $M$  is a sphere.

If  $n > d+3$  then all the  $i$ -faces occur in  $M$  for  $i \leq n-d-3 \geq 1$ . Therefore the link of any vertex in  $M$  is an  $(n-1)$ -vertex combinatorial  $(d-1)$ -sphere with  $f$ -vector satisfying:  $f_i = \binom{n-1}{i+1}$  for  $0 \leq i \leq n-d-4$ . Hence by (6), the  $h$ -vector of this link satisfies  $h_i = \binom{n-d-2+i}{i}$  for  $0 \leq i \leq n-d-3$ .

If  $d = 2c$  then by (7) for  $n > 3c+3$ , we get  $\binom{n-c-3}{c-1} = h_{c-1} = h_{c+1} = \binom{n-c-1}{c+1}$ . Which gives  $n = 2c+2$ , contrary to our assumption in this case.

If  $d = 2c-1$  then for  $n \geq 3c+3$ , we get  $\binom{n-c-2}{c-1} = h_{c-1} = h_c = \binom{n-c-1}{c}$ . Which gives  $n = 2c+1$ , a contradiction.

Thus,  $n \leq 3c+3$  if  $d$  is even and  $n < 3c+3$  if  $d$  is odd. Therefore, by (BK1) and (BK2)  $M$  is either a sphere or a "manifold like a projective plane". But as Euler characteristic of  $M$  is odd,  $M$  cannot be a sphere. This completes the proof.

## Acknowledgement

The author is thankful to B Bagchi for suggesting this problem and for numerous useful conversations. This work has been done when the author was a Visiting Scientist at the Indian Statistical Institute, Bangalore, and the author expresses his gratitude for their hospitality and support. The author is also thankful to the referee for pointing out the fact that complementarity implies the Euler characteristic is odd, which helped to shorten the proof.

## References

- [1] Arnoux P and Marin A, The Kühnel triangulation of complex projective plane from the view-point of complex crystallography (part II), *Memoirs of the Faculty of Sc., Kyushu Univ. Ser. A* 45 (1991) 167–244
- [2] Bagchi B and Datta B, On Kühnel's 9-vertex complex projective plane, *Geometriae Dedicata* (to appear)
- [3] Brehm U and Kühnel W, Combinatorial manifolds with few vertices, *Topology* 26 (1987) 467–473
- [4] Brehm U and Kühnel W, 15-vertex triangulation of an 8-manifold, *Math. Ann.* 294 (1992) 167–193
- [5] Eells Jr J and Kuiper N H, Manifolds which are like projective plane, *Publ. Math. I.H.E.S.* 14 (1962) 181–222
- [6] Kühnel W and Banchoff T F, The 9-vertex complex projective plane, *The Math. Intell.* 5 No. 3 (1983) 11–22
- [7] Kühnel W and Laßmann G, The unique 3-neighbourly 4-manifold with few vertices, *J. Combin. Theory (A)* 35 (1983) 173–184
- [8] Stanley R P, The Upper Bound Conjecture and Cohen-Macaulay Rings, *Stud. Appl. Math.* LIV (1975) 135–142
- [9] Stanley R P, The Number of Faces of a Simplicial Convex Polytope, *Adv. Math.* 35 (1980) 236–238

## Deformations of complex structures on $\Gamma \backslash SL_2(\mathbb{C})$

C S RAJAN

School of Mathematics, Tata Institute of Fundamental Research, Homi Bhabha Road,  
 Bombay 400 005, India

MS received 21 July 1993

**Abstract.** Let  $G$  be a connected complex semisimple Lie group. Let  $\Gamma$  be a cocompact lattice in  $G$ . In this paper, we show that when  $G$  is  $SL_2(\mathbb{C})$ , nontrivial deformations of the canonical complex structure on  $X$  exist if and only if the first Betti number of the lattice  $\Gamma$  is non-zero. It may be remarked that for a wide class of arithmetic groups  $\Gamma$ , one can find a subgroup  $\Gamma'$  of finite index in  $\Gamma$ , such that  $\Gamma'/[\Gamma', \Gamma']$  is finite (it is a conjecture of Thurston that this is true for all cocompact lattices in  $SL(2, \mathbb{C})$ ).

We also show that  $G$  acts trivially on the coherent cohomology groups  $H^i(\Gamma \backslash G, \mathcal{O})$  for any  $i \geq 0$ .

**Keywords.** Deformations; lattice; cohomology.

### 1. Introduction

Let  $M$  be a compact smooth manifold. We assume that  $M$  can be equipped with a hyperbolic structure. In ([3]), Johnson and Millson show that the space of deformations of ‘marked conformal structures’ on  $M$  has dimension at least  $r$ , where  $r$  is the largest number of disjoint, nonsingular, totally geodesic hypersurfaces in  $M$ . Such hypersurfaces are known to contribute to the first Betti number of  $M$ . Now, it is known that if the dimension of  $M$  is  $n$ , then  $M$  is diffeomorphic to  $\Gamma \backslash SO(n, 1)/K$ , where  $\Gamma$  is a torsion-free cocompact lattice in  $SO(n, 1)$  and  $K$  is a maximal compact subgroup of  $SO(n, 1)$ . When  $n = 3$ ,  $SO(3, 1)$  is locally isomorphic to  $SL(2, \mathbb{C})$  and thus carries a complex structure. One can raise the question, whether there exists nontrivial deformations of the complex structure on  $\Gamma \backslash SL(2, \mathbb{C})$ , and if so whether the deformations are related to the ‘topology of  $\Gamma$ ’.

In a different direction, Matsushima raised the question whether the canonical complex structure on  $\Gamma \backslash G$  is infinitesimally rigid, where  $G$  is a connected complex semisimple Lie group and  $\Gamma$  is an irreducible torsion-free cocompact lattice in  $G$ . In [8], Raghunathan showed that whenever  $G$  has no 3-dimensional components, the canonical complex structure on  $\Gamma \backslash G$  is infinitesimally rigid. It is easy to extend this result to all  $G$ , provided  $G$  is not three-dimensional. We remark that when  $G$  is not three dimensional, the first Betti number of  $\Gamma$  is zero.

From these results, we are thus led to relating the ‘topology of  $\Gamma$ ’ to the deformations of the complex structure on  $\Gamma \backslash G$ , where  $G = SL(2, \mathbb{C})$ . Our main result states that nontrivial deformations of the canonical complex structure on  $\Gamma \backslash SL(2, \mathbb{C})$  exist if and only if the first Betti number of the cocompact torsion-free lattice  $\Gamma$  in  $SL(2, \mathbb{C})$  is nonzero.

and a sublattice of rank index in  $\Gamma$ , such that the first Dehn number of  $\Gamma$  is nonzero. Thurston's conjecture is known to be true for a wide class of arithmetic lattices  $\Gamma$  ([5], [6], [7]).

We also show that the natural action of  $G$  on  $H^*(\Gamma \backslash G, \mathcal{O})$  is the trivial action for  $G$  a complex semisimple Lie group and  $\Gamma$  any cocompact torsion-free lattice in  $G$ .

## 2. Cohomology computations

Let  $G$  be a connected, semisimple complex Lie group with Lie algebra  $L(G)$  of left invariant vector fields on  $G$ . Let  $L(G)^{\mathbb{C}}$  denote the complexification of  $L(G)$ . Let  $\Gamma$  be a discrete subgroup of  $G$  and let  $X = \Gamma \backslash G$ . Elements of  $L(G)^{\mathbb{C}}$  can be regarded as complex vector fields on  $X$ . Then the space  $U_1$  of holomorphic left invariant vector fields on  $G$  project to  $X$ , to give a trivialisation of the holomorphic tangent bundle of  $X$ . Let  $\mathcal{O}$  denote the sheaf of germs of holomorphic functions on  $X$ . There is a natural holomorphic action of  $G$  on  $H^*(X, \mathcal{O})$ . Our aim in this section is to show that  $G$  acts trivially on  $H^*(\Gamma \backslash G, \mathcal{O})$ .

Let  $K$  be a maximal compact subgroup of  $G$ . Let  $Y = \Gamma \backslash G/K$  and let  $\pi$  denote the natural projection  $X \rightarrow Y$ . Let  $L(K)$  be the Lie algebra of  $K$ .

### PROPOSITION 1

*Let  $\rho$  be a nontrivial holomorphic representation of  $G$  on  $F$ . Then for any  $i \geq 0$ ,*

$$H^i(\Gamma, \rho) = (0)$$

*Proof.* By Matsushima's formula it is enough to show the following:

$$H^i(L(G), L(K); W \otimes F) = (0) \quad \forall i \geq 0,$$

where  $W$  is an irreducible unitary  $(L(G), L(K))$ -module (see [2, page 224]).

Let  $L(H)^+ \subset L(K)$  be a maximal abelian subalgebra of  $L(K)$ . Let  $L(H)$  be the centralizer in  $L(G)$  of  $L(H)^+$ . Then  $L(H)$  is a Cartan subalgebra of  $L(G)$ . Let  $\Phi$  denote the root system of  $(L(G)^{\mathbb{C}}, L(H)^{\mathbb{C}})$ . Let  $\theta$  denote the Cartan involution of the pair  $(L(G), L(K))$ . Fix a positive system of roots  $\Phi^+$  of  $\Phi$  as in ([2, page 65]). In particular this implies that if  $\alpha$  is a positive root, so is  $\theta\alpha$ . Let  $\delta$  denote half the sum of the positive roots. Then  $\theta\delta = \delta$ .

Let  $J$  denote the complex structure on  $L(G)$ . We can choose  $\theta$  such that  $\theta J \theta = -J$ . The space of holomorphic (resp. antiholomorphic) vectors of  $L(G)^{\mathbb{C}}$  can be taken as the kernel of  $(J - i)$  (resp.  $(J + i)$ ). By the compatibility relation between  $\theta$  and  $J$  given above,  $\theta$  interchanges the space of holomorphic and antiholomorphic vectors of  $L(G)^{\mathbb{C}}$ .

Let  $\lambda$  be the highest weight of  $F^*$  with respect to the ordering defined by  $\Phi^+$ . Since the representation  $\rho$  is assumed to be holomorphic,  $\lambda$  vanishes on the space of antiholomorphic vectors of  $L(H)^{\mathbb{C}}$ . But then  $\theta\lambda$  vanishes on the space of holomorphic vectors of  $L(H)^{\mathbb{C}}$ . Since  $\lambda$  is assumed to be nontrivial we have  $\theta\lambda \neq \lambda$ . Since  $\theta\delta = \delta$ , we have  $\theta(\lambda + \delta) \neq \lambda + \delta$ . But then by Proposition 6.12 1), page 69 of ([2]),  $H^*(L(G), L(K), W \otimes F) = (0)$ . Hence the proposition.

*Remark.* For the purposes of studying the deformations of  $\Gamma \backslash SL_2(\mathbb{C})$ , it is enough to have this proposition when  $G = SL_2(\mathbb{C})$ . In this case, the proposition has been proved by Raghunathan. See [2, page 225].

Let  $L_\rho$  be the local system on  $X$  associated to the representation  $\rho$ .

In ([8]), it is shown that the  $E_2$  term of the Hodge-de Rham spectral sequence, which converges to  $H^*(X, L_\rho)$  is given by

$$E_2^{pq} = H^p(L(G), H^q(X, \mathcal{O}) \otimes_{\mathbb{C}} F),$$

where the  $G$ -module structure on  $H^q(X, \mathcal{O}) \otimes_{\mathbb{C}} F$  is the tensor product of the representations.

There is also the Leray spectral sequence associated to the  $K$ -principal fibration  $\pi: X \rightarrow Y$ , converging to  $H^*(X, L_\rho)$ , and whose  $E_2$  term is given by

$$'E_2^{pq} = H^p(Y, R^q \pi_* L_\rho).$$

Now for a fibration  $\pi: X \rightarrow Y$  with fiber  $Z$ , and given a local system  $L_\rho$  on  $X$ ,  $R^q \pi_* L_\rho$  is the local system associated to the representation of  $\pi_1(Y)$  on  $H^q(Z, L_\rho|_Z)$ .

In the above situation, since  $\Gamma$  is torsion-free,  $L_\rho|_K$  is a trivial local system and hence  $R^q \pi_* L_\rho = H^q(K, \mathbb{C}) \otimes_{\mathbb{C}} F$  as  $\pi_1(Y) = \Gamma$ -modules, where the action of  $\Gamma$  on  $H^q(K, \mathbb{C})$  is trivial and on  $F$  it acts as  $\rho$ . Thus,

$$'E_2^{pq} = H^p(\Gamma, \rho) \otimes_{\mathbb{C}} H^q(K, \mathbb{C})$$

With notation as above, we prove the following:

**Theorem 1.**  $G$  acts trivially on  $H^i(\Gamma \backslash G, \mathcal{O})$  for all  $i \geq 0$ .

*Proof.* Let  $\rho$  denote an irreducible, holomorphic representation of  $G$  on  $F$ , occurring in  $\text{Hom}(H^i(X, \mathcal{O}), \mathbb{C})$ . Suppose  $\rho$  is nontrivial. Then by the proposition proved above,  $H^p(\Gamma, \rho) = (0) \forall p \geq 0$ . Hence  $'E_2^{pq} = 0$  and so we have that  $H^p(X, L_\rho) = (0) \forall p \geq 0$ .

We have by the Hodge-de Rham spectral sequence computed above

$$E_2^{pq} = H^p(L(G), H^q(X, \mathcal{O}) \otimes_{\mathbb{C}} F)$$

and this converges to  $H^*(X, L_\rho)$ .

We claim that  $E_2^{pq} = 0 \forall p, q \geq 0$ . If not, there is a maximal  $p_0$  and  $q_0$  such that  $E_2^{pq} \neq 0$ , i.e.,  $E_2^{pq} = (0)$  if either  $p > p_0$  or  $q > q_0$  and  $E_2^{p_0 q_0} \neq (0)$ . Since the differentials  $d_r (r \geq 2)$  of the spectral sequence increases the indices  $p$  or  $q$  by at least 1, we see that  $E_2^{p_0 q_0}$  survives to  $E_\infty$  and is nonvanishing. But this contradicts the vanishing of  $H^*(X, L_\rho)$  shown above. Hence  $E_2^{pq} = (0) \forall p, q \geq 0$ .

But then  $E_2^{i0} = H^0(L(G), H^i(X, \mathcal{O}) \otimes_{\mathbb{C}} F)$  is nonzero since by assumption on  $F$ ,  $H^i(X, \mathcal{O}) \otimes_{\mathbb{C}} F$  contains a copy of the trivial representation and hence the invariants  $H^0(L(G), H^i(X, \mathcal{O}) \otimes_{\mathbb{C}} F)$  can never be zero. Hence  $\rho$  has to be trivial and this proves the theorem.

## PROPOSITION 2

*Proof.* Since  $X = \Gamma \backslash G$  is compact,  $H^*(X, \mathcal{O})$  are finite dimensional. By Whitehead lemma for semisimple Lie algebras,  $H^1(L(G), \mathbb{C}) = H^2(L(G), \mathbb{C}) = (0)$  (see [2]). Hence in the Hodge-de Rham spectral sequence given above (for  $F = \mathbb{C}$ ), we have  $E_2^{20} = E_2^{20} = 0$ .

Therefore  $E_\infty^{10} = 0$  and  $E_\infty^{01} = E_2^{01}$ . Hence  $H^1(X, L_\rho) = H^0(L(G), H^1(X, \mathcal{O}))$ . But from the Leray spectral sequence given above, we have  $'E_2^{01} = 0$  and  $'E_\infty^{10} = 'E_2^{10} = H^1(\Gamma, \mathbb{C})$ . Hence  $H^1(\Gamma, \mathbb{C}) = H^1(X, L_\rho)$ . Since  $L(G)$  acts trivially on  $H^1(X, \mathcal{O})$  by the theorem proved above, we have

$$H^1(\Gamma, \mathbb{C}) = H^1(X, L_\rho) = H^0(L(G), H^1(X, \mathcal{O})) = H^1(X, \mathcal{O}).$$

This proves the proposition.

*Remark.* When  $G$  has more than one three dimensional component and  $\Gamma$  is an irreducible lattice in  $G$ , then the first Betti number of  $\Gamma$  vanishes by a theorem of Bernstein-Kazhdan, see ([1].) Let  $\Theta$  denote the sheaf of germs of holomorphic vector fields on  $X$ . Since the holomorphic tangent bundle of  $X$  is trivial,

$$H^1(X, \Theta) = H^1(X, \mathcal{O}) \otimes U_1.$$

By the above proposition  $H^1(X, \Theta)$  vanishes. This extends the following rigidity theorem of Raghunathan for groups  $G$  with no three dimensional components, (see [8]). (note that it is enough to prove the results for simply connected  $G$ ).

**Theorem 2.** *Let  $G$  be a connected complex semisimple Lie group and  $\Gamma$  an irreducible cocompact lattice in  $G$ . Assume that  $G$  is not locally isomorphic to  $SL(2, \mathbb{C})$ . Then the canonical complex structure on  $X$  is rigid.*

Essentially the only interesting case left is when  $G = SL_2(\mathbb{C})$ . In this case if the canonical complex structure on  $X$  is not locally rigid, then  $H^1(X, \Theta) \neq 0$ . Hence we obtain that the first Betti number of  $\Gamma$  is nonzero.

### 3. Deformations of $\Gamma \backslash SL_2(\mathbb{C})$

From now onwards  $G = SL_2(\mathbb{C})$ .

In this section we show the existence of non-trivial deformations of the canonical complex structure on  $X$ , whenever the first Betti number of  $\Gamma$  is nonzero.

Let  $T$  denote the holomorphic tangent bundle of  $X$ . By means of the projection  $G \rightarrow X$ ,  $U_1$  defines a trivialisation of  $T$ . Let  $A^p$  denote the space of  $(0, p)$  forms of  $X$  with values in  $T$ . We have  $\bar{\partial}: A^p \rightarrow A^{p+1}$ . It is well known that on  $A = \sum_{p \geq 0} A^p$ , one can define a bilinear operation,

$$[,] : A^p \otimes A^q \rightarrow A^{p+q}$$

which turns  $A$  into a graded Lie algebra complex. The graded Lie algebra structure descends down to a graded Lie algebra structure on  $H = \sum_{p \geq 0} H^{(0,p)}(T)$ . The group  $G$  acts as graded Lie algebra automorphisms on  $A$  and this action descends to an action of  $G$  on  $H$ , compatible with the graded Lie algebra structure on  $H$ .

Let  $K$  be a maximal compact subgroup of  $G$ . Fix a hermitian metric on  $T$ , on which  $K$  acts as isometries.

With respect to this metric one can define the adjoint  $\bar{\partial}^*$  of  $\bar{\partial}$ , the Laplacian  $\Delta = \bar{\partial}\bar{\partial}^* + \bar{\partial}^*\bar{\partial}$ , the Green's operator  $G$  and the harmonic projection operator  $H$ . Since  $K$  acts as isometries the action of  $K$  commutes with that of  $\Delta$ ,  $G$  and  $H$ .

By means of the Dolbeault isomorphism  $H^p(X, \Theta) = H^{0,p}(T)$  and by the theory of harmonic forms, we can think of  $H^p(X, \Theta)$  as a subspace of  $A^p$  consisting of harmonic forms and these spaces are isomorphic as  $K$ -modules.

**Theorem 3.** *Let  $G = SL_2(\mathbb{C})$  and  $\Gamma$  a cocompact lattice on  $G$ . Let  $X = \Gamma \backslash G$ . Assume that the first Betti number of  $\Gamma$  is nonzero. Then there exists nontrivial deformations of the canonical complex structure on  $X$ .*

*Proof.* Let  $\Theta$  denote the sheaf of germs of holomorphic vector fields on  $X$ . Since  $T$  is trivialised by the projection of  $U_1$  to  $X$ , we have as  $G$ -modules,

$$H^1(X, \Theta) = H^1(X, \mathcal{O}) \otimes U_1$$

and

$$H^2(X, \Theta) = H^2(X, \mathcal{O}) \otimes U_1$$

where the  $G$ -module structure is the tensor product of the  $G$ -modules. Since the duality relationship is compatible with the  $G$ -action and the canonical bundle is trivial  $H^1(X, \Theta) \simeq H^2(X, \Theta)$  as  $G$ -modules.

Fix a maximal torus  $S$  of  $K$  and a system of positive roots of  $G$  with respect to  $S$  and let  $\lambda$  be a highest weight for the representation of  $K$  on  $H^1(X, \Theta)$ . Let  $V$  denote the corresponding highest weight subspace of  $H^1(X, \Theta)$ .  $V$  is nonzero as  $H^1(X, \Theta)$  is a nontrivial  $L(G)$  module. We also think of  $H^1(X, \Theta)$  as the subspace of harmonic forms in  $A^1$ .

Even though  $H^1(X, \Theta) \simeq H^2(X, \Theta) (\neq 0)$ , we will show below that the method of Kodaira–Spencer as outlined in ([4] p. 316), can be adapted to the subspace  $V$  of  $H^1(X, \Theta)$ , to obtain the existence of nontrivial deformations of the complex structure on  $X$ . Let  $\{\beta_1, \dots, \beta_m\}$  be a basis of  $V$  and put

$$\phi_1(t) = \beta_1 t_1 + \dots + \beta_m t_m.$$

Define inductively a sequence  $\{\phi_k(t)\}_{k \geq 2}$  of  $A^1$  valued homogeneous polynomials of degree  $k$  in the variables  $t_1, \dots, t_m$  as follows:

$$\phi_k(t) = \frac{1}{2} \sum_{l=1}^{k-1} \bar{\partial}^* G[\phi_l(t), \phi_{k-l}(t)]$$

Let  $\phi(t) = \sum_{k=1}^{\infty} \phi_k(t)$ . The inductive definition of the  $\phi_k$  is made so as to secure the condition

$$\phi(t) = \phi_1(t) + \frac{1}{2} \bar{\partial}^* G[\phi(t), \phi(t)]$$

$\phi(t)$  defines an almost complex structure on the real submanifold underlying the space  $X$ . One can proceed as in ([4], p. 316), to show that  $\phi(t)$  converges with respect to the Holder norm for  $|t| < \varepsilon$ , provided  $\varepsilon > 0$  is sufficiently small and that  $\phi(t)$  is  $C^\infty$  on  $X \times \Delta_\varepsilon$ , where  $\Delta_\varepsilon = \{t \in \mathbb{C}^m : |t| < \varepsilon\}$ .

The almost complex structure given by  $\phi(t)$  is integrable if

$$\mathbf{H}[\phi(t), \phi(t)] = 0$$

(see Lemma 6.3 [4], p. 316).

Now  $[\cdot, \cdot]: A^1 \otimes A^1 \rightarrow A^2$  is a morphism of  $G$ -modules and if  $\omega, \eta \in A^1$  are of weights  $k\lambda$  and  $l\lambda$  respectively for the compact torus  $S$ , then  $[\omega, \eta]$  is of weight  $(k+l)\lambda$ .  $\phi_1(t)$  is of weight  $\lambda$  with respect to the maximal torus  $S$  in  $K$ . By induction one can check that  $[\phi_l(t), \phi_{k-l}(t)]$  is of weight  $k\lambda$  for the torus  $S$  ( $0 < l < k$ ). Since the action of  $K$  commutes with  $\bar{\partial}^*$  and  $G$ , we see that  $\phi_k(t)$  is of weight  $k\lambda$  for the torus  $S \subset K$ . Hence  $[\phi(t), \phi(t)]$  is a sum of elements of weight  $k\lambda$ , where  $k \geq 2$ . Since  $\text{Ker } \Delta$  has highest weight  $\lambda$ , and the action of  $K$  commutes with  $\Delta$  and  $\mathbf{H}$  we see that

$$\mathbf{H}[\phi(t), \phi(t)] = 0.$$

As in ([4]), one can also check that this forms a complex analytic family on  $\Delta_e$ , such that the Kodaira–Spencer deformation map from the tangent space of the base space at 0 to  $H^1(X, \Theta)$  maps  $T(\Delta_e)_0$  isomorphically to  $V$ . Hence one gets a nontrivial family of deformations parametrized by  $V$  of the complex manifold  $X$ .

*Remark.* Let  $C$  be the cone of all highest weight vectors in  $H^1(X, \Theta)$ . By what we have shown above, we see that  $C$  is a complex analytic subvariety of the base space parametrizing the Kuranishi family. Hence we get a complex analytic family in the sense of Kuranishi of deformations of the complex structure on  $X$  over  $C$ , which is a subfamily of the universal family Kuranishi has constructed. It is not known whether this family is complete.

*Remark.* Thurston's conjecture states that given a uniform lattice  $\Gamma$  in  $SL(2, \mathbb{C})$ , there is a sublattice whose first Betti number is nonzero. We have shown that this conjecture is equivalent to showing existence of nontrivial deformations of the complex structure of some suitable finite cover of  $X$ . For a wide class of arithmetic lattices the conjecture has been shown to be true ([5], [6], [7]).

*Jump phenomenon.*  $G$  acts on  $C$  and the highest weight vectors in the same  $G$  orbit give rise to identical complex structures on the smooth manifold underlying  $X$ .

If we take a highest weight vector  $v$ , then the Borel subgroup in  $G$  corresponding to  $v$ , acts by scaling. Hence we have that the complex manifolds  $X_t$  are all isomorphic for  $t \neq 0$  and not isomorphic to the complex manifold  $X = X_0$ .

### Acknowledgement

The author wishes to express his sincere gratitude to M S Raghunathan who suggested this problem and provided him with invaluable guidance.

### References



- [2] Borel A and Wallach N C, Continuous cohomology, discrete subgroups and representations of reductive groups, *Ann. Math. Stud.* (Princeton: Univ. Press) **94** (1980)
- [3] Johnson D and Millson J J, Deformation spaces associated to compact hyperbolic manifolds in *Discrete Groups in Geometry and Analysis*; Proceedings of a Conference held at Yale University in honor of G. D. Mostow *Progress in Mathematics Series* (Birkhauser) ed. R Howe (1985)
- [4] Kodaira K, *Complex manifolds and deformations of complex structure* (Berlin: Springer-Verlag) (1986)
- [5] Lebesse J P and Schwermer J, On liftings and cusp cohomology of arithmetic groups, *Invent. Math.* **83** (1986) 383–401
- [6] Millson J J, On the first Betti number of a constant negatively curved manifold, *Ann. Math.* **104** (1976) 235–247
- [7] Millson J J and Raghunathan M S, Geometric construction of cohomology for arithmetic groups, *Proc. Indian Acad. Sci. (Math. Sci.)* **90** (1981) 103–123
- [8] Raghunathan M S, Vanishing theorems for cohomology groups associated to discrete subgroups of semisimple groups, *Osaka J. Math.* **67** (1966) 243–256



# Differential subordinations concerning starlike functions

S PONNUSAMY

School of Mathematics, SPIC Science Foundation, 92, G N Chetty Road, Madras 600 017, India

MS received 25 April 1992; revised 22 February 1993

**Abstract.** Denote by  $S^*(\rho)$ , ( $0 \leq \rho < 1$ ), the family consisting of functions  $f(z) = z + a_2 z^2 + \dots + a_n z^n + \dots$  that are analytic and starlike of order  $\rho$ , in the unit disc  $|z| < 1$ . In the present article among other things, with very simple conditions on  $\mu$ ,  $\rho$  and  $h(z)$  we prove the  $f''(z)(f(z)/z)^{\mu-1} \prec h(z)$  implies  $f \in S^*(\rho)$ . Our results in this direction then admit new applications in the study of univalent functions. In many cases these results considerably extend the earlier works of Miller and Mocanu [6] and others.

**Keywords.** Differential subordination; univalent; starlike and convex functions.

## 1. Introduction

Let  $\mathcal{A}$  denote the class of functions  $f$  that are analytic in the unit disc  $\Delta$  and  $f(0) = f'(0) - 1 = 0$ , with  $S^*(\rho)$ ,  $0 \leq \rho < 1$ , designating the subclass of  $\mathcal{A}$  consisting of functions starlike (univalent) of order  $\rho$ . We denote by  $\mathcal{P}$  the class of functions  $p$  that are analytic in  $\Delta$  so that  $p(0) = 1$ .

From a result of Libera [5], it is known that if  $p \in \mathcal{P}$  and if  $\lambda$  is a function defined on  $\Delta$  with  $\operatorname{Re} \lambda(z) > 0$  for  $z \in \Delta$ , then

$$p(z) + \lambda(z)zp'(z) \prec \frac{1+z}{1-z} \text{ implies } p(z) \prec \frac{1+z}{1-z}, \quad z \in \Delta, \quad (1)$$

where  $\prec$  denotes subordination. This is an example of *non-autonomous differential subordination*; that is we allow functions of  $z$  to be present in addition to the terms  $p(z)$  and  $zp'(z)$ . However the above result has been improved in [8] when  $\operatorname{Re} \lambda(z) > \eta > 0$ ,  $z \in \Delta$  and the sharp estimation has been obtained in [9] only when  $\lambda(z) \equiv \alpha$ ,  $\operatorname{Re} \alpha \geq 0$ .

Using the Herglotz' representation theorem for functions with positive real part, one easily obtains that if  $p$  is analytic in  $\Delta$  and  $\alpha > 0$ , then

$$\frac{1+z}{1-z} - \alpha \frac{z}{(1-z)^2} p(z) \prec \frac{1+z}{1-z} \text{ implies } p(z) \prec \frac{1+z}{1-z}, \quad z \in \Delta. \quad (2)$$

This provides another example of *non-autonomous differential subordination*.

In [7, Theorem 5], Miller and Mocanu determined conditions on  $\alpha$  and  $\beta$ , for which

$$p(z) + \lambda(z)zp'(z) < \left(\frac{1+z}{1-z}\right)^\alpha \text{ implies } p(z) < \left(\frac{1+z}{1-z}\right)^\beta, \quad z \in \Delta, \quad (3)$$

only when  $\lambda(z) \equiv 1$ . Clearly (1), (2) and (3) are not sharp, in general.

In this note we first extend the above result by determining a general condition on the function  $\lambda(z)$ . Secondly we determine conditions on  $\lambda \in \mathbb{C}$ ,  $0 \leq \rho < 1$  and  $\mu > 0$  to obtain

$$f'(z) \left( \frac{f(z)}{z} \right)^{\mu-1} < 1 + \lambda z \text{ implies } f \in S^*(\rho), \quad z \in \Delta.$$

Then we generate this implication to obtain new results concerning Libera–Bernardi integral operators.

## 2. Preliminaries

We shall need the following definitions and Lemmas:

Let  $\mathcal{H}$  denote the class of analytic functions in  $\Delta$ .

A function  $f \in \mathcal{H}$  is said to belong to  $B(\alpha, \rho)$  if  $f$  satisfies the differential equation

$$f'(z) \left( \frac{f(z)}{z} \right)^{\alpha-1} = H(z)$$

where  $H$  is of the form

$$H(z) = h(z) \left( \frac{g(z)}{z} \right)^{\alpha-1}, \quad h(0) = 1,$$

with  $\alpha \geq 0$ ,  $zg'(z)/g(z) < (1+z)/(1-z)$  and  $\operatorname{Re} h(z) > \rho$ . For  $0 \leq \rho < 1$ , the functions in  $B(\alpha, \rho)$  are in fact univalent in  $\Delta$  and is a subclass of the well-known class of Bazilevič functions of the type  $(\alpha, 0)$  [1]. Furthermore we denote by  $B_1(\alpha, \rho)$  the subclass of  $B(\alpha, \rho)$  for which  $g(z) = z$ ; and let  $S^*(\rho) \equiv B_1(0, \rho)$ .

It is interesting to observe that the function  $f$  defined by

$$f(z)/z = ((1+z)/(1-z))^{1/\alpha}$$

is not in  $B_1(\alpha, \rho)$ , for  $0 \leq \rho < 1$ , since

$$f'(z) \left( \frac{f(z)}{z} \right)^{\alpha-1} = \frac{1+z}{1-z} + \frac{2z}{\alpha(1-z)^2} \rightarrow -1/\alpha \text{ as } z \rightarrow i;$$

on the other hand  $f(z) = z$  is in  $B_1(\alpha, \rho)$  for each  $\rho < 1$ .

*Lemma A.* Let  $\Omega$  be a set in  $\mathbb{C}$  and let  $q$  be analytic and univalent on  $\bar{\Delta}$  except for those  $\zeta \in \partial\Delta$  for which  $\lim_{z \rightarrow \zeta} q(z) = \infty$ . Suppose that  $\psi: \mathbb{C}^2 \times \Delta \rightarrow \mathbb{C}$  satisfies the condition

then  $p(z) \prec q(z)$  in  $\Delta$ .

If  $q(z) = (1+z)/(1-z)$ , then the condition (4) simplifies to

$$\psi(ix, y; z) \notin \Omega, \quad \text{for } z \in \Delta \text{ and reals } x, y \text{ with } y \leq -(1+x^2)/2. \quad (5)$$

Substituting  $q(z) = Jz$ , we see that  $q(\zeta) = Je^{i\theta}$  and  $\zeta q'(\zeta) = Je^{i\theta}$ . Then the condition (4) simplifies to

$$\psi(Je^{i\theta}, Me^{i\theta}; z) \notin \Omega \text{ for } z \in \Delta \quad (5')$$

when

$$M = mJ \geq J \text{ and } \theta \text{ real.}$$

**Lemma B.** Let  $f$  be analytic in  $\Delta$  and let  $g$  be analytic and univalent in  $\bar{\Delta}$ , with  $f(0) = g(0)$ . If  $f$  is not subordinate to  $g$ , then there exist points  $z_0 \in \Delta$  and  $\zeta_0 \in \partial\Delta$ , and an  $m \geq 1$  for which

$$(a) \quad f(|z| < |z_0|) \subset g(|z| < |z_0|),$$

$$(b) \quad f(z_0) = g(\zeta_0), \text{ and}$$

$$(c) \quad z_0 f'(z_0) = m \zeta_0 g'(\zeta_0).$$

**Lemma C.** For  $0 \leq \theta \leq \pi$  and  $-1 < \beta \leq \bar{\beta} \approx 4.567$ ,

$$\frac{1}{1+\beta} + \sum_{k=1}^n \frac{\cos k\theta}{k+\beta} \geq 0.$$

**Lemma D.** If  $p \in \mathcal{P}$  and  $\text{Re } p(z) > 1/2$  in  $\Delta$ , then for any function  $f$ , analytic in  $\Delta$ , the function  $p * f$  takes values in the convex hull of the image of  $\Delta$  under  $f$ .

Lemmas A and B are due to Miller and Mocanu [6, 7] and Lemma C is due to Gasper [4]. The widely used assertion of Lemma D readily follows by using Herglotz' representation theorem for  $p$ .

### 3. Main results

**Lemma 1.** Let  $\beta_0$  be the solution of

$$\beta\pi = 3\pi/2 - \tan^{-1}(\eta), \quad (6)$$

for a suitable fixed  $\eta > 0$  so that  $\lambda(z): \Delta \rightarrow \mathbb{C}$  satisfies

$$|\text{Im } \lambda(z)| \leq \frac{1}{\eta} \left( \text{Re } \lambda(z) - \frac{\eta}{\beta} \right), \quad z \in \Delta \quad (7)$$

and let

$$\alpha = \alpha(\beta, \eta) = \beta + (2/\pi) \tan^{-1}(\eta), \quad 0 < \beta \leq \beta_0. \quad (8)$$

If  $p \in \mathcal{P}$ , then

$$|\arg(p(z) + \lambda(z)zp'(z))| < \frac{\pi}{2}\alpha, \quad z \in \Delta \quad (9)$$

implies  $|\arg p(z)| < \frac{\pi}{2}\beta$ ,  $z \in \Delta$ .

For specific values of  $\beta$  and  $\eta$  Lemma 1 simplifies. In particular if  $p \in \mathcal{P}$ , then we present a pair of interesting examples illustrating Lemma 1:

*Example 1.* If we fix  $\beta = 2/3$  and  $\eta = 1/\sqrt{3}$  in the above Lemma we obtain  $\alpha(2/3, 1/\sqrt{3}) = 1$  and so

$$\operatorname{Re}(p(z) + \lambda(z)zp'(z)) > 0 \text{ implies } |\arg p(z)| < \pi/3, \quad z \in \Delta$$

provided  $|\operatorname{Im} \lambda(z)| \leq \sqrt{3}(\operatorname{Re} \lambda(z) - \sqrt{3}/2)$  in  $\Delta$ .

*Example 2.* If we take  $\beta = 1$  and  $\eta = \tan((\alpha - 1)\pi/2)$ , Lemma 1 yields

$$p(z) + \lambda(z)zp'(z) < \left(\frac{1+z}{1-z}\right)^\alpha \text{ implies } \operatorname{Re} p(z) > 0, \quad z \in \Delta$$

when  $|\operatorname{Im} \lambda(z)| \leq \cot(\pi(\alpha - 1)/2)[\operatorname{Re} \lambda(z) - \tan(\pi(\alpha - 1)/2)]$  in  $\Delta$ .

In particular we easily have the following:

(i) if  $\lambda$  satisfies the condition that  $|\operatorname{Im} \lambda(z)| \leq (\operatorname{Re} \lambda(z) - 1)$  in  $\Delta$ , then

$$|\arg(p(z) + \lambda(z)zp'(z))| < 3\pi/4 \text{ implies } \operatorname{Re} p(z) > 0, \quad z \in \Delta;$$

(ii) if  $0 \leq \gamma < 1$ , then

$$p(z) + \tan(\gamma\pi/2)zp'(z) < \left(\frac{1+z}{1-z}\right)^{\gamma+1} \text{ implies } \operatorname{Re} p(z) > 0, \quad z \in \Delta.$$

*Proof of Lemma 1.* Let  $H_r$  be defined by

$$H_r(z) = [(1+z)/(1-z)]^r, \quad z \in \Delta, \quad (0 < r < 2).$$

Then

$$H_r(\Delta) \subset \left\{ \omega \in \mathbb{C} : |\arg \omega| < \frac{\pi r}{2} \right\}$$

and therefore (9) is equivalent to

$$p(z) + \lambda(z)zp'(z) \in H_\alpha(\Delta).$$

We need to prove that  $p(z) < H_\beta(z)$ .

Assume that  $p$  is not subordinate to  $H_\beta(z)$ . By Lemma B, there exist point  $z_0 \in \Delta$  and  $\zeta_0 \in \partial\Delta$  that satisfy (a) to (c). Let us first consider the case  $p(z_0) \neq 0$  (and hence

$\zeta_0 \neq \pm 1$ ). Using the conditions (a) to (c) of Lemma B, together with

$$ix = \frac{1 + \zeta_0}{1 - \zeta_0}, \quad (x - \text{real}),$$

we obtain

$$p(z_0) + \lambda(z_0)z_0 p'(z_0) = H_\beta(\zeta_0) + m\lambda(z_0)\zeta_0 H'_\beta(\zeta_0) \equiv R(x, m), \text{ say}$$

where

$$R(x, m) = (ix)^\beta [1 + im\beta \frac{1}{2}(x + \frac{1}{x})\lambda(z_0)].$$

Now in this case, it suffices to show that

$$R(x, m) \notin H_\alpha(\Delta), \quad (10)$$

for  $z_0 \in \Delta$  and all real  $x \neq 0$  and  $m \geq 1$ .

For convenience we let

$$\lambda(z_0) = U + iV$$

and

$$\Phi = \frac{m\beta U}{(2|x|/(1+x^2)) + m\beta|V|}.$$

Then we may write

$$\arg R(x, m) = \pm \beta \frac{\pi}{2} + \tan^{-1}(M(x, m)) \quad (11)$$

where + sign is for  $x > 0$  and - sign is for  $x < 0$ , and

$$M(x, m) = \frac{m\beta U}{(2x/(1+x^2)) - m\beta V}.$$

First we note that for all  $m \geq 1$  and  $x \neq 0$ ,

$$\begin{aligned} \Phi &\geq \frac{m\beta U}{1 + m\beta|V|} \\ &\geq \frac{\beta U}{1 + \beta|V|} \\ &\geq \eta, \quad (\text{by (7)}), \end{aligned} \quad (12)$$

and  $\tan^{-1}$  is an increasing function. We also note that if  $x \neq 0$  and  $m\beta V = 2x/(x^2 + 1)$  then

$$\arg R(x, m) = \begin{cases} \frac{\beta(\pi + 1)}{2} & \text{if } x > 0 \\ -\frac{\beta(\pi + 1)}{2} & \text{if } x < 0. \end{cases}$$

Now, to complete the proof, we consider the following six cases:

- (i)  $x > 0$  and  $V \leq 0$
- (ii)  $x > 0$  and  $0 < Vm\beta < 2x/(1+x^2)$
- (iii)  $x > 0$  and  $Vm\beta > 2x/(1+x^2)$
- (iv)  $x < 0$  and  $V \geq 0$
- (v)  $x < 0$  and  $0 > Vm\beta > 2x/(1+x^2)$
- (vi)  $x < 0$  and  $Vm\beta < 2x/(1+x^2)$ .

Therefore it is elementary to see that

$$M(x, m) \begin{cases} = \Phi & \text{when case (i) holds;} \\ \geq \Phi & \text{when case (ii) holds;} \\ \leq -\Phi & \text{when case (iii) holds;} \\ = -\Phi & \text{when case (iv) and (v) hold;} \\ \geq \Phi & \text{when case (vi) holds.} \end{cases} \quad (13)$$

For cases (i) and (ii), by (12) and (13), we get

$$\begin{aligned} \arg R(x, m) &\geq \frac{\beta\pi}{2} + \tan^{-1}(\Phi) \\ &\geq \frac{\beta\pi}{2} + \tan^{-1}(\eta) \end{aligned}$$

and for cases (iv) and (v), we obtain

$$\begin{aligned} \arg R(x, m) &\geq \frac{-\beta\pi}{2} + \tan^{-1}(-\Phi) \\ &\geq \frac{-\beta\pi}{2} - \tan^{-1}(\eta). \end{aligned}$$

For case (iii), we estimate

$$\begin{aligned} \arg R(x, m) &\leq \frac{\beta\pi}{2} - \tan^{-1}(\eta) \\ &= \frac{\beta\pi}{2} - (\alpha - \beta)\frac{\pi}{2}, \quad \text{by (8),} \\ &< 2\pi - \frac{\alpha\pi}{2}, \quad \text{since } 0 < \beta \leq \beta_0 < 2 \end{aligned}$$

and for case (vi), we similarly have

$$\begin{aligned} \arg R(x, m) &\geq \frac{-\beta\pi}{2} + \tan^{-1}(\eta) \\ &= -\beta\pi + \frac{\alpha\pi}{2}, \quad \text{by (8),} \end{aligned}$$



$$> -\left[2\pi - \frac{\alpha\pi}{2}\right], \quad \text{since } 0 < \beta \leq \beta_0 < 2.$$

Therefore,

$$2\pi - \alpha\pi/2 \geq |\arg R(x, m)| \geq \alpha\pi/2$$

for all  $0 \neq x$  real and  $m \geq 1$ . This implies (10) and since it contradicts the hypothesis, we must have  $p < H_\beta$  in  $\Delta$ . Using a similar argument that of [7, Theorem 5], the case  $p(z_0) = 0$  for which  $0 \notin H_\beta(\Delta)$  (possible only when  $\beta \geq 1$ ) can be handled easily. This completes the proof of Lemma 1.  $\square \square$

As an immediate consequence of Lemma 1, we have

**Theorem 1.** Suppose that  $\alpha'$ , ( $1/2 < \alpha' < 1$ ), is the unique root of the equation

$$\mu = (1 - \alpha) \cot\left(\frac{2\alpha - 1}{2} \pi\right) \quad (14)$$

for a suitable fixed  $\mu > 0$ . Then for  $f \in \mathcal{A}$ ,

$$\left| \arg f'(z) \left( \frac{f(z)}{z} \right)^{\mu-1} \right| < \frac{\alpha' \pi}{2} \text{ implies } f \in S^*(0), \quad z \in \Delta.$$

*Proof.* Since  $1/2 < \alpha' < 1$ ,  $f$  is univalent in  $\Delta$ . By (14) we get

$$\alpha = (1 - \alpha) + \frac{2}{\pi} \tan^{-1} \left( \frac{1 - \alpha}{\mu} \right)$$

which is equivalent to (8) with  $\beta = 1 - \alpha$  and  $\lambda = 1/\mu = \eta$ .

Since

$$\left| \arg \left( \frac{zf'(z)}{f(z)} \right) \right| \leq \left| \arg f'(z) \left( \frac{f(z)}{z} \right)^{\mu-1} \right| + \left| \arg \left( \frac{f(z)}{z} \right)^\mu \right|,$$

the conclusion now follows from Lemma 1 (for  $\lambda = 1/\mu = \eta$  and  $p(z) = (f(z)/z)^\mu$ ) using the hypothesis of Theorem 1.  $\square \square$

If we choose  $\mu = 1$  in Theorem 1, we have

**Example 3.** For  $f \in \mathcal{A}$ ,

$$|\arg f'(z)| < \alpha' \pi/2 \text{ implies } f \in S^*(0), \quad z \in \Delta,$$

provided  $\alpha' \approx 0.616 \dots$ .

A proof very similar to that of [6, Theorem 5] gives the following. We just give a simple proof because of its independent interest.

**Lemma 2.** Let  $B(z)$  and  $C(z)$  be functions defined on  $\Delta$  with  $B(z) \neq 0$  satisfying either

$$|B(z) + C(z)| \geq D/J \text{ and } \operatorname{Re}[C(z)/B(z)] \geq -1, \quad z \in \Delta,$$

or,

$$|\operatorname{Im}(C(z)/B(z))| \geq D/(|B|J), \quad z \in \Delta.$$

If  $p$  is analytic in  $\Delta$ , with  $p(0) = 0$ , and if

$$|C(z)p(z) + B(z)zp'(z)| < D, \quad z \in \Delta \quad (15)$$

then  $|p(z)| < J$  in  $\Delta$ .

*Remark 1.* Now we construct two examples to show that most of the special cases of Lemma 2 may be stated as best possible. These follow by taking different choices for  $C(z)$  and  $B(z)$ .

*Example 4.* Let  $p \in \mathcal{H}$ . Then for  $k$ , a complex number satisfying  $\operatorname{Re} k \geq -|k|^2$ , we have

$$|p(z) + kzp'(z)| < J \text{ implies } |p(z)| < J/|k+1|, \quad z \in \Delta. \quad (16)$$

It is clear that the estimation in (16) is sharp.

Upon taking  $p(z) = (f(z)/zf'(z)) - 1$ , (16) easily deduces the following: If  $f \in \mathcal{A}$  with  $f'(z)f(z)/z \neq 0$  in  $\Delta$  and  $\lambda \in \mathbb{C}$ ,  $0 < |\lambda| < |k+1|$ , then we have

$$\left| (k-1) \left( \frac{zf'(z)}{f(z)} - 1 \right) + k \left( \frac{zf''(z)}{f'(z)} \right) \right| < |\lambda| \left| \frac{zf'(z)}{f(z)} \right| \text{ implies } \frac{zf'(z)}{f(z)} < \frac{k+1}{k+1+\lambda z}.$$

This for  $k=1$  yields the well-known result of Robertson [11].

*Example 5.* Let  $\phi \in \mathcal{A}$  with  $\phi(z)\phi'(z) \neq 0$  for  $z \neq 0$ . Then for  $f \in \mathcal{H}$ , we obtain

$$|f(z)| < D \text{ implies } \left| z^{-k+1} \phi(z)^{-1} \int_0^z f(t) t^{k-1} \phi'(t) dt \right| < J, \quad z \in \Delta \quad (17)$$

provided  $\phi$  satisfies

$$\left| 1 + k \frac{\phi(z)}{z\phi'(z)} \right| \geq \frac{D}{J} \text{ and } \operatorname{Re} \left( k + \frac{z\phi'(z)}{\phi(z)} \right) \geq 0, \quad z \in \Delta. \quad (18)$$

If  $k$  is real and  $k > 0$ , ( $\eta > 0$ ), the condition (18) will hold if

$$\operatorname{Re} \left( \frac{\phi(z)}{z\phi'(z)} \right) \geq \frac{\eta - k}{2} \text{ and } D^2 \leq J^2 \left[ 1 + \eta \left| \frac{\phi(z)}{z\phi'(z)} \right|^2 \right], \quad z \in \Delta$$

and hence this gives (17).

A special case of this result leads to the following: if  $\phi \in \mathcal{A}$ , with  $\phi(z)\phi'(z) \neq 0$  for  $z \neq 0$ , satisfies

$$\operatorname{Re} \frac{z\phi'(z)}{\phi(z)} \geq \frac{\eta - 1}{2} \text{ and } D = \left[ J^2 + \eta \max_{z \in \Delta} \left| \frac{\phi(z)}{z\phi'(z)} \right|^2 \right]^{1/2},$$

then

$$|f(z)| < D \text{ implies } \left| \phi(z)^{-1} \int_0^z f(t) \phi'(t) dt \right| < J, \quad z \in \Delta. \quad (19)$$

Furthermore if we assume  $\phi(z) = z^n$ , ( $n = 1, 2, 3, \dots$ ) then (19) yields that

$$|f(z)| < [J^2 + (2n+1)/n^2]^{1/2} \text{ implies } \left| nz^{-n} \int_0^z f(t) t^{n-1} dt \right| \leq |z|J, \quad z \in \Delta.$$

This improves the result of Miller and Mocanu [6, p. 211].

*Proof of Lemma 2.* If we let  $\psi(r, s; z) = C(z)r + B(z)s$ , then (15) becomes

$$|\psi(p(z), zp'(z); z)| < D \text{ in } \Delta.$$

Now we use (5'). For  $M \geq J$  and  $z \in \Delta$  we obtain

$$\begin{aligned} |\psi(Je^{i\theta}, Me^{i\theta}; z)|^2 - D^2 &= |B|^2 M^2 + 2J \operatorname{Re}(C\bar{B})M + |C|^2 J^2 - D^2 \\ &= |B|^2 \left[ (M + J \operatorname{Re}(C/B))^2 - \left\{ J^2 (\operatorname{Re}(C/B))^2 - \frac{|C|^2 J^2 - D^2}{|B|^2} \right\} \right]. \end{aligned}$$

Since  $M \geq J$ , the hypothesis of Lemma 2 now implies that

$$|\psi(Je^{i\theta}, Me^{i\theta}; z)| \geq D$$

for  $z$  in  $\Delta$ . The conclusion of this Lemma now follows by applying Lemma A with  $q(z) = Jz$ .  $\square \square$

*Lemma 3.* Let  $\lambda$  be a complex number,  $\mu > 0$  and let  $Q$  be a function defined on  $\Delta$  satisfying the condition

$$Q(z) < \frac{\lambda\mu}{\mu+1} z, \quad z \in \Delta \quad (20)$$

so that

$$\beta \equiv \beta(\mu, |\lambda|) = \frac{\mu+1 - |\lambda|\sqrt{(\mu+1)^2 + \mu^2}}{\mu+1 + |\lambda|} \text{ and } |\lambda| \leq \frac{\mu+1}{\sqrt{(\mu+1)^2 + \mu^2}}. \quad (21)$$

If  $p \in \mathcal{P}$  satisfies

$$Q(z)(\beta + (1-\beta)p(z)) + (1-\beta)(p(z)-1) < \lambda z, \quad z \in \Delta \quad (22)$$

then  $\operatorname{Re} p(z) > 0$  in  $\Delta$ .

*Proof.* Let  $\psi(r; z) = Q(z)(\beta + (1-\beta)r) + (1-\beta)(r-1)$ . Then (22) becomes

$$|\psi(p(z); z)| < |\lambda|, \quad z \in \Delta. \quad (23)$$

Now if we set  $Q(z) = U + iV$ , then for this  $\psi$  and  $x$ -real we must have

$$\begin{aligned} |\psi(ix; z)|^2 &= [(U+1)\beta - 1 - V(1-\beta)x]^2 + [(1-\beta)(U+1)x + V\beta]^2 \\ &= (1-\beta)^2 |U+1 + iV|^2 x^2 + 2(1-\beta)Vx + |\beta(U+iV) - (1-\beta)|^2. \end{aligned}$$

Hence for all real  $x$  and  $z \in \Delta$

$$|\psi(ix; z)| \geq |\lambda| \quad (24)$$

holds if  $Q$  satisfies

$$V^2 \leq |U + 1 + iV|^2 [|\beta(U + iV) - (1 - \beta)|^2 - |\lambda|^2]. \quad (25)$$

By (20) and (21), we note that

$$\begin{aligned} |\beta(U + iV) - (1 - \beta)| &\geq (1 - \beta) - \beta|U + iV| \\ &> (1 - \beta) - \beta|\lambda|(\mu/(\mu + 1)) \\ &= [|\lambda|\sqrt{(\mu + 1)^2 + \mu^2}]/(\mu + 1). \end{aligned}$$

This shows that

$$|\beta(U + iV) - (1 - \beta)|^2 - |\lambda|^2 > |\lambda|^2 \mu^2 / (\mu + 1)^2.$$

In view of this inequality, we see that (25) will follow if we can show that

$$|V| \leq |\lambda||U + 1 + iV|\mu/(1 + \mu);$$

or, equivalently,

$$|\operatorname{Im} Q(z)| \leq |\lambda|[\operatorname{Re} Q(z) + 1]\mu/[(\mu + 1)^2 - |\lambda|^2 \mu^2]^{1/2}, \quad z \in \Delta. \quad (26)$$

However from (20), a little computation shows that (26) is satisfied; i.e. under the hypotheses of Lemma 3, (24) holds for all real  $x$  and all  $z \in \Delta$ . Therefore, by (23) and Lemma A with  $q(z) = (1 + z)/(1 - z)$  and  $\Omega = \{\omega \in \mathbb{C} : |\omega| < |\lambda|\}$ , we obtain  $\operatorname{Re} p(z) > 0$  in  $\Delta$ . This completes the proof.  $\square \square$

**Theorem 2.** Let  $\beta$  and  $\lambda$  be given by (21) and  $\mu > 0$ . If  $f \in \mathcal{A}$  satisfies

$$f'(z) \left( \frac{f(z)}{z} \right)^{\mu-1} < 1 + \lambda z, \quad z \in \Delta \quad (27)$$

then  $f \in S^*(\beta)$ .

*Proof.* Since  $|\lambda| < 1$  and since  $f$  satisfies (27),  $f$  is univalent in  $\Delta$ . Therefore as an application of (16) with  $p(z) = (f(z)/z)^\mu$ , we have

$$\left( \frac{f(z)}{z} \right)^\mu < 1 + \frac{\mu\lambda}{\mu + 1} z, \quad z \in \Delta. \quad (28)$$

If we consider

$$\frac{zf'(z)}{f(z)} = \beta + (1 - \beta)p(z)$$

and

$$Q(z) = \left( \frac{f(z)}{z} \right)^\mu - 1$$

then from (27) and (28) it is easy to see that

$$f'(z) \left( \frac{f(z)}{z} \right)^{\mu-1} - 1 = Q(z)(\beta + (1-\beta)p(z)) + (1-\beta)(p(z)-1) < \lambda z.$$

The desired conclusion now follows from Lemma 3.  $\square \square$

If we take  $\mu = 1$  in the above theorem we have the following:

### COROLLARY 1

For  $f \in \mathcal{A}$  and  $\lambda \leq 2/\sqrt{5}$ , we have

$$|f'(z) - 1| < \lambda \text{ implies } f \in S^* \left( \frac{2 - \sqrt{5}\lambda}{2 + \lambda} \right), \quad z \in \Delta.$$

*Remark 2.* Observe that the order of starlikeness increases from 0 to 1 as  $\lambda$  decreases from  $2/\sqrt{5}$  to 0.

A special case of Corollary 1 (i.e. when  $\lambda = 2/\sqrt{5}$ ) was obtained in [12]. In [2, Theorem 2], Fournier showed that the condition  $0 < \lambda \leq 2/\sqrt{5}$  is actually necessary and sufficient for the special case to hold. However our method of proof is not only different but allows us to find similar estimate for a more general problem as in Theorem 2. In his later work Fournier [3, Corollary 1] continued his effort in obtaining the order of starlikeness of function  $f \in \mathcal{A}$  with  $|f'(z) - 1| < \lambda$  in  $\Delta$ . Using function theoretic approach this particular result has been further improved by Ponnusamy and Singh [10] in a different context. There is no other reason in restricting  $\lambda$  in such a way that  $0 < \lambda \leq 2/\sqrt{5}$ , but to avoid the use of negative order of starlikeness in the conclusion of Corollary 1.

*Lemma 4.* If  $f \in \mathcal{A}$  satisfies

$$|f'(z) + \alpha z f''(z) - 1| < \lambda, \quad z \in \Delta, \quad (29)$$

then for  $\alpha \geq \alpha' > 0$ ,

$$|f'(z) + \alpha' z f''(z) - 1| < \frac{\alpha' + 1}{\alpha + 1} \lambda, \quad z \in \Delta.$$

*Proof.* Consider the identity

$$f'(z) + \alpha' z f''(z) - 1 = \left( \frac{\alpha - \alpha'}{\alpha} \right) [f'(z) - 1] + \frac{\alpha'}{\alpha} [f'(z) + \alpha z f''(z) - 1]. \quad (30)$$

Thus by (16), (29) yields

$$f'(z) < 1 + \frac{\lambda}{\alpha + 1} z, \quad z \in \Delta. \quad (31)$$

Since  $0 < \alpha'/\alpha \leq 1$ , (31) and (30) give

$$|f'(z) + \alpha' z f''(z) - 1| < \left( \frac{\alpha - \alpha'}{\alpha} \right) \frac{\lambda}{\alpha + 1} + \frac{\alpha'}{\alpha} \lambda = \frac{\alpha' + 1}{\alpha + 1} \lambda$$

and that ends the proof of Lemma 4.  $\square \square$

*Example 6.* Let  $R(\alpha, J) = \left\{ f \in \mathcal{A} : \left| (1 - \alpha) \frac{f(z)}{z} + \alpha f'(z) - 1 \right| < J, \quad z \in \Delta \right\}$ . Now from (16) it follows that if  $\alpha > -1$  and if  $f \in R(\alpha, J)$ , then

$$\left| \frac{f(z)}{z} - 1 \right| < \frac{J}{\alpha + 1}, \quad z \in \Delta.$$

From the analysis similar to that used for the proof of Lemma 4, we conclude that

$$f \in R(\alpha, J) \text{ implies } |f'(z) - 1| < \frac{2J}{\alpha + 1}, \quad z \in \Delta;$$

i.e., the functions in  $R(\alpha, J)$  are in fact close-to-convex (with respect to  $e^{i\phi}z$ , for  $\phi$ -real) and bounded in  $\Delta$  when  $\alpha \geq 2J - 1$ . Furthermore from Corollary 1, we get

$$f \in R(\alpha, J) \text{ implies } f \in S^* \left( \frac{\alpha + 1 - J\sqrt{5}}{\alpha + 1 + J} \right)$$

provided  $\alpha \geq J\sqrt{5} - 1$ .

*Example 7.* Let  $f(z) = z + \sum_{k=2}^{\infty} a_k z^k$  be in  $\mathcal{A}$  and set  $f_n(z) = z + \sum_{k=2}^n a_k z^k$  be its  $n$ th section. Then, according to Lemma D, the function  $L$  defined by

$$L(z) = \frac{1}{2} + \frac{1}{1 + \beta} + \sum_{k=2}^n \frac{k[1 + (k-1)\alpha]}{k + \beta - 1} a_k z^{k-1}, \quad (-1 < \beta \leq \bar{\beta} \approx 4.567), \quad z \in \Delta$$

takes values in the convex hull of the image of  $\Delta$  under  $G(z) = f'(z) + \alpha z f''(z)$ , since

$$L(z) = \left( \frac{1}{2} + \frac{1}{1 + \beta} + \sum_{k=2}^n \frac{z^{k-1}}{k + \beta - 1} \right) * G(z)$$

in which bracketed term in the above identity, by Lemma C, lies in  $\{\omega \in \mathbb{C} : \operatorname{Re} \omega > 1/2\}$  for each  $z$  in  $\Delta$ .

In particular this observation for  $\beta = 1$  given that if  $G$ , defined by  $G(z) = f'(z) + \alpha z f''(z)$ , satisfies some property then so does the function  $(1 - \alpha)(f_n(z)/z) + \alpha f'_n(z)$ .

**Theorem 3.** Let  $\alpha$  be a complex number such that  $\operatorname{Re} \alpha > 0$  and  $|\alpha + 1| \geq \sqrt{5}\gamma/2$ . If  $f \in \mathcal{A}$  satisfies

$$|f'(z) + \alpha z f''(z) - 1| < \gamma, \quad z \in \Delta$$

then

$$f \in S^* \left( \frac{2|\alpha + 1| - \sqrt{5}\gamma}{2|\alpha + 1| + \sqrt{5}\gamma} \right)$$

Corollary 1 by taking  $\lambda = \gamma/(\alpha + 1)$ .  $\square \square$

As an example of the above theorem we obtain: if  $f \in \mathcal{A}$  and  $\alpha \geq (\sqrt{5}\gamma - 2/2)$ , then

$$|f'(z) + \alpha z f''(z) - 1| < \gamma \text{ implies } f \in S^* \left( \frac{2(\alpha + 1) - \sqrt{5}\gamma}{2(\alpha + 1) + \gamma} \right);$$

and therefore if  $\gamma = 1$ , this leads to

$$|f'(z) + \alpha z f''(z) - 1| < 1 \text{ implies } f \in S^* \left( \frac{2\alpha + 2 - \sqrt{5}}{2\alpha + 3} \right) \text{ provided } \alpha \geq \frac{\sqrt{5} - 2}{2}.$$

## COROLLARY 2

If  $\mu > 0$ ,  $c > -\mu$ , and  $0 < K \leq ((\mu + c + 1)(\mu + 1))/(\mu + c)[(\mu + 1)^2 + \mu^2]^{1/2}$  and if  $f \in \mathcal{A}$  satisfies

$$\left| f'(z) \left( \frac{f(z)}{z} \right)^{\mu-1} - 1 \right| < K, \quad z \in \Delta$$

then the function  $F = I_{\mu, c}(f)$ , defined by

$$F(z) = \left[ \frac{\mu + c}{z^c} \int_0^z f^\mu(t) t^{c-1} dt \right]^{1/\mu}, \quad z \in \Delta, \quad (32)$$

for  $F(z)/z \neq 0$  in  $\Delta$ , belongs to  $S^*(\beta')$ , where  $\beta'$  is defined by

$$\beta'(\mu, c, K) = \frac{(\mu + 1)(\mu + c + 1) - K(\mu + c)[(\mu + 1)^2 + \mu^2]^{1/2}}{(\mu + 1)(\mu + c + 1) + K(\mu + c)}. \quad (33)$$

*Proof.* Consider the function  $P$  defined by

$$P(z) = F'(z) \left( \frac{F(z)}{z} \right)^{\mu-1}, \quad z \in \Delta.$$

Then from (32) we easily get

$$P(z) + \frac{1}{\mu + c} z P'(z) = f'(z) \left( \frac{f(z)}{z} \right)^{\mu-1}, \quad (34)$$

and hence by (16) and the hypotheses, (34) yields

$$P(z) = F'(z) \left( \frac{F(z)}{z} \right)^{\mu-1} < 1 + \lambda z, \quad z \in \Delta,$$

with

$$|\lambda| = \frac{K(\mu + c)}{\mu + c + 1} \leq (\mu + 1)/\sqrt{(\mu + 1)^2 + \mu^2}.$$

This shows that all the required conditions of Theorem 2 are satisfied for  $F$ . Therefore we conclude that  $F \in S^*(\beta')$ , where  $\beta'$  is given by (33).  $\square \square$

Corresponding to the case  $\mu = 1$  (and  $c = 0$  resp.) of the above Corollary we have the following:

*Example 8.* Let  $f \in \mathcal{A}$ . Then

$$|f'(z) - 1| < K \text{ implies } I_{1,c}(f) \in S^* \left( \frac{2 - [K(1+c)\sqrt{5/(2+c)}]}{2 + [K(1+c)/(2+c)]} \right), \quad z \in \Delta, \quad (35)$$

provided  $c > -1$  and  $0 < K \leq 2(c+2)/[\sqrt{5}(c+1)]$ .

For instance, if we let  $K = 1$  and  $c = 1$  in (35), and if  $f \in \mathcal{A}$  then for Libera transform, we have

$$|f'(z) - 1| < 1 \text{ implies } \frac{2}{z} \int_0^z f(t) dt \in S^*((3 - \sqrt{5})/4), \quad z \in \Delta.$$

*Example 9.* Let  $f \in \mathcal{A}$ . Then

$$|f'(z) - 1| < K \text{ implies } I_{\mu,0}(f) \in S^* \left( \frac{(\mu+1)^2 - K\mu[(\mu+1)^2 + \mu^2]^{1/2}}{(\mu+1)^2 + K_\mu} \right), \quad (36)$$

provided  $\mu > 0$  and  $0 < K \leq (\mu+1)^2/(\mu\sqrt{(\mu+1)^2 + \mu^2})$ .

In particular for  $\mu = 1$ , (36) reduces to

$$f \in \mathcal{A} \text{ and } |f'(z) - 1| < K \text{ implies } \int_0^z \frac{f(t)}{t} dt \in S^* \left( \frac{4 - K\sqrt{5}}{4 + K} \right),$$

for  $0 < K \leq 4/\sqrt{5}$ .

## Acknowledgement

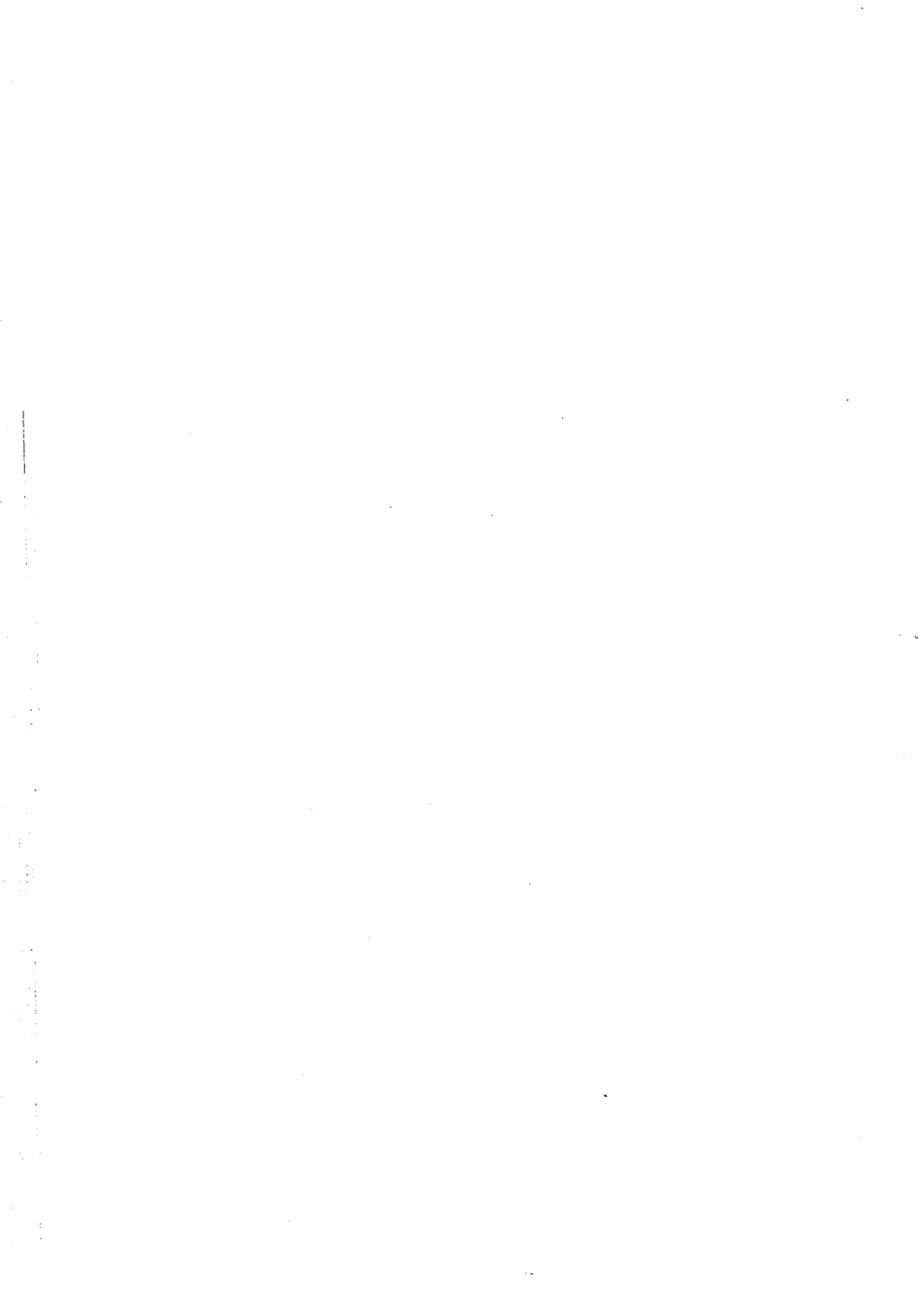
The author is grateful to Prof. V Singh in helping him to know the results of [2, 3, 12] and thanks the referee of the paper for his useful comments. This work is supported by National Board for higher mathematics.

## References

- [1] Bazilevič I E, On a case of integrability in quadratures of Loewner-Kurfarev equation, *Mat. Sb.* **37** (1955) 471-476
- [2] Fournier R, On integrals of bounded analytic functions in the closed unit disc, *Complex Variables* **11** (1989) 125-133
- [3] Fournier R, The range of a continuous linear functional over a class of functions defined by



- [5] Libera R J, Some classes of regular univalent functions, *Proc. Am. Math. Soc.* **16** (1965) 755–758
- [6] Miller S S and Mocanu P T, Differential subordination and inequalities in the complex plane, *J. Differ. Equ.* **67** (1987) 199–211
- [7] Miller S S and Mocanu P T, Marx-Strohhäcker differential subordination systems, *Proc. Am. Math. Soc.* **99** (1987) 527–534
- [8] Ponnusamy S and Karunakaran V, Differential subordination and conformal mappings, *Complex Variables* **11** (1989) 79–86
- [9] Ponnusamy S, Differential subordination and starlike functions, *Complex Variables* **19** (1992), 185–194
- [10] Ponnusamy S and Singh V, Convolution properties of some classes of analytic functions, (submitted)
- [11] Robertson M S, Certain classes of starlike functions, *Michigan Math. J.* **32** (1985) 135–140
- [12] Singh V, Univalent functions with bounded derivative in the unit disc, *Indian J. Pure Appl. Math.* **8** (1977) 1370–1377



## On the structure of stable random walks

JON AARONSON

School of Mathematical Sciences, Tel Aviv University, 69978 Tel Aviv, Israel

MS received 9 June 1993; revised 28 September 1993

**Abstract.** We show that the Cauchy random walk on the line, and the Gaussian random walk on the plane are similar as infinite measure preserving transformations.

**Keywords.** Stable random walks; Gaussian random walk.

### 1. Introduction

A measure preserving transformation  $T$  is considered acting on a *standard* measure space  $(X_T, B_T, m_T)$  (a complete, separable metric space equipped with its Borel sets and a  $\sigma$ -finite, non-atomic measure). It is known that standardness is unaffected by replacing  $X_T$  with a  $T$ -invariant subset  $X'_T \in B_T$  of full measure, and we shall consider  $T$  acting on  $(X_T, B_T, m_T)$  to be the same as  $T$  acting on  $(X'_T, B_T \cap X'_T, m_T)$ .

Let  $S$  and  $T$  be measure preserving transformations. A *factor map* from  $S$  to  $T$  is a map  $\pi: X_S \rightarrow X_T$  such that

$$\pi S = T\pi, \quad \pi^{-1}B_T \subset B_S, \quad \text{and} \quad m_S \circ \pi^{-1} = c m_T$$

where  $0 < c < \infty$ .

In this situation (denoted by  $\pi: S \rightarrow T$ ), one says that  $T$  is a *factor* of  $S$  and that  $S$  is an extension of  $T$  (both denoted  $S \rightarrow T$ ).

It is necessary to consider factor maps with  $c \neq 1$  as our measure spaces are not normalized. The constant  $c$  can be thought of as a relative normalization of the transformations concerned.

Two measure preserving transformations are said to be *similar* if they have a common extension, that is: if there is another measure preserving transformation of which they are both factors; and they are said to be *strongly disjoint* if they have no common extension. We denote the statement that  $S$  and  $T$  are similar by  $S \sim T$ .

Any two transformations preserving finite measures are similar, their Cartesian product being a common extension. Invariants for similarity are given in [A1, A2], where it is shown that similarity is an equivalence relation. Examples of conservative, ergodic, measure preserving transformations which are strongly disjoint from their inverses are given in [A2].

In this paper, we consider random walks on  $\mathbb{R}$  and  $\mathbb{R}^2$ . For  $f$  a probability on  $G$ , a locally compact second countable abelian group, the *random walk* on  $G$  with *jump distribution*  $f$  can be defined as follows: Let  $S_f$  be the shift on  $G^{\mathbb{Z}}$  considered with the  $S_f$ -invariant product measure  $m_f = \Pi f$ . The *random walk* on  $G$  with *jump distribution*

$f$  is the measure preserving transformation  $T_f$  defined on  $(G^{\mathbb{Z}} \times G, m_f \times m_G)$  (where  $m_G$  is Haar measure on  $G$ ) by

$$T_f(x, y) = (S_f x, y + x_0).$$

The structure of random walks on  $\mathbb{Z}$  and  $\mathbb{Z}^2$  has been considered in [AK] and [ALP] where conditions for isomorphism are given.

For  $\alpha \in [1, 2]$  let  $f_\alpha$  denote the symmetric  $\alpha$ -stable law on  $\mathbb{R}$  with characteristic function

$$\int_{\mathbb{R}} e^{itx} df_\alpha(x) = e^{-|t|^\alpha}.$$

It is well known that  $f_1$  has a Cauchy density,  $f_2$  has a Gaussian density, and indeed,  $f_\alpha$  is absolutely continuous with strictly positive, continuous density  $\forall \alpha \in [1, 2]$ . Let  $T_\alpha = T_{f_\alpha}$ , the random walk on  $\mathbb{R}$  with  $\alpha$ -stable jump distribution.

**Theorem.**

$$T_1 \sim T_2 \times T_2.$$

*Remark.* For  $\alpha, \alpha' \in [1, 2]$  not both 2, we have that  $(1/\alpha + 1/\alpha') > 1$ , and  $T_\alpha \times T_{\alpha'}$  is dissipative, and isomorphic to  $z \mapsto z + 1$  on  $\mathbb{R}^2$  equipped with Lebesgue measure. This will be established below.

The method of proof of the theorem is by renewal theory. In §2 we formulate, and deduce the theorem from the main lemma, (also proving the remark). The main lemma also shows (via [A1]) that the transformations  $\{T_\alpha: \alpha \in [1, 2]\}$  are pairwise strongly disjoint.

## 2. Renewal theory: the main lemma

Recall from [Fe] that a bounded sequence of non-negative real numbers  $u = (u_0, u_1, \dots)$  is called a *renewal sequence* if  $u_0 = 1$  and there is a sequence of non-negative real numbers  $g = g(u) = (g_1, g_2, \dots)$  satisfying the *renewal equation*

$$u_n = \sum_{k=1}^n g_k u_{n-k} \quad \forall n \in \mathbb{N}.$$

The renewal sequence  $u$  is called *recurrent* if  $g(u)$  is a probability on  $\mathbb{N}$  (that is  $\sum_{n=1}^{\infty} g_n = 1$ ).

Let  $T$  be a conservative ergodic measure preserving transformation. Recall from [A1] that a set  $A \in \mathcal{B}_T$ ,  $0 < m_T(A) < \infty$  is called a *recurrent event* for  $T$  if, for every  $0 = n_0 < n_1 < \dots < n_k$ ,

$$m_T\left(\bigcup_{k=0}^K T^{-n_k} A \mid A\right) = \prod_{k=1}^K u_{n_k - n_{k-1}}$$

where  $u_n = m_T(T^{-n} A \mid A)$ .

$$u = u(A) := (u_0, u_1, \dots)$$

defined by  $u_n = m_T(T^{-n}A|A)$  is a recurrent renewal sequence. Conversely, every recurrent renewal sequence corresponds in the above manner to a recurrent event of some conservative ergodic measure preserving transformation. Let  $u$  be a recurrent renewal sequence, and let  $g = g(u)$  be the associated probability on  $\mathbb{N}$  satisfying the renewal equation.

One can define ([Ch]) a stochastic matrix  $P = P_u$  with state space  $\mathbb{N}$  by

$$p_{j,k} = \begin{cases} g_k & \text{if } j = 1 \\ 1 & \text{if } j - k = 1, \\ 0 & \text{else.} \end{cases}$$

This matrix is irreducible, recurrent, and has the stationary distribution

$$m_k = \sum_{j=k}^{\infty} g_j.$$

Let  $T_u$  denote the Markov shift of  $(P_u, m)$ , that is the shift on  $\mathbb{N}^{\mathbb{Z}}$  equipped with the  $T_u$ -invariant measure  $\mu$  defined by

$$\mu([s_1, \dots, s_n]_k) = m_{s_1} p_{s_1, s_2} \dots p_{s_{n-1}, s_n}$$

where  $[s_1, \dots, s_n]_k = \{x \in \mathbb{N}^{\mathbb{Z}} : x_{k+j} = s_j \forall 1 \leq j \leq n\}$ , then  $T_u$  is a conservative ergodic measure preserving transformation, and the set  $[1]_0$  is a recurrent event for  $T_u$  with renewal sequence  $u$ .

It was shown in [A1] that if  $A$  is a recurrent event for (the conservative ergodic measure preserving transformation)  $T$ , then  $T \rightarrow T_{u(A)}$ .

If  $u, u'$  are recurrent renewal sequences, then  $uu'$  (defined by  $(uu')_n = u_n u'_n$ ) is a renewal sequence, and if  $uu'$  is recurrent, then

$$T_u \times T_{u'} \rightarrow T_{uu'}.$$

This is because  $T_u \times T_{u'}$  is the Markov shift of  $P_u \times P_{u'}$ .

For  $\beta > 0$ , define  $u(\beta)$  by

$$u_n(\beta) = \left( \frac{1}{n+1} \right)^{\beta} \quad (n \geq 0),$$

then  $u_{n+1}/u_n \uparrow$  as  $n \uparrow$ , so by Kaluza's theorem ([Kal, Kin]),  $u(\beta)$  is a renewal sequence, which is recurrent iff  $\beta \leq 1$ . Note that  $u(\beta)u(\beta') = u(\beta + \beta')$ . We are now in a position to state the

*Main Lemma.*

$$T_{\alpha} \sim T_{u(1/\alpha)} \quad \forall \alpha \in [1, 2].$$

Given the main lemma, the theorem follows easily:

$$T_1 \sim T_{u(1)} = T_{u(1/2)u(1/2)} \leftarrow T_{u(1/2)} \times T_{u(1/2)} \sim T_2 \times T_2.$$

The truth of the remark is also established easily. Write  $(1/\alpha) + (1/\alpha') = 1 + 2\varepsilon$  where  $\varepsilon > 0$ , then

$$T_\alpha \times T_{\alpha'} \sim T_{u(1/\alpha)} \times T_{u(1/\alpha')} \leftarrow T_{u(1/\alpha)} \times T_{u(1/\alpha' - \varepsilon)} \times T_{u(\varepsilon)} := S.$$

The transformation  $T_{u(1/\alpha)} \times T_{u(1/\alpha' - \varepsilon)}$  is dissipative. If  $W$  is a generating wandering set for it, then  $W \times X_{T_{u(\varepsilon)}}$  is an infinite generating wandering set for  $S$ . It follows that  $T_\alpha \times T_{\alpha'}$  (being similar to  $S$ ) also has an infinite generating wandering set, and is therefore isomorphic to  $z \mapsto z + 1$  on  $\mathbb{R}^2$  considered with respect to Lebesgue measure.

The rest of this paper is therefore devoted to the proof of the main lemma, which is in two stages.

The first stage is to show that there is a recurrent renewal sequence  $w(\alpha)$  such that  $T_\alpha \sim T_{w(\alpha)}$ . This is a mild restatement of [AN].

The second stage is to show that  $T_{w(\alpha)} \sim T_{u(1/\alpha)}$ . This uses the notion of equivalence of renewal sequences introduced in [AK]. Recall from [AK] that two renewal sequences  $u$ , and  $u'$  are *equivalent* (denoted  $u \sim u'$ ) if there are positively recurrent, aperiodic renewal sequences  $v$ , and  $v'$  such that  $uv = u'v' = w$ . In this situation,

$$T_u \leftarrow T_u \times T_v \rightarrow T_w$$

whence  $T_u \sim T_{u'}$  when  $u \sim u'$ .

### 3. Proof of the main lemma

Fix  $\alpha \in [1, 2]$  and consider the positive operator  $P$  defined on  $L^1(\mathbb{R})$  by

$$P_g(x) := \int_{\mathbb{R}} g(x+y) df_\alpha(y).$$

**Lemma 1.** ([AN]) *There exists  $q \in (0, 1)$  such that if*

$$w_0 = 1, w_n = q \int_I P^n 1_I dm \quad (n \geq 1),$$

*then  $w$  is a renewal sequence, and  $T_\alpha \sim T_w$ . Here,  $m = m_{\mathbb{R}}$  denotes Lebesgue measure on  $\mathbb{R}$ , and  $I = [0, 1]$ .*

**Lemma 2.** *Let  $w$  be as in lemma 1, then*

$$w \sim u(1/\alpha).$$

Clearly, the main lemma follows from lemmas 1 and 2.

**Proof of Lemma 1.** The transformation  $T_\alpha$  is isomorphic to  $T_P$ , the *shift* of  $P$ , which is the shift on  $\mathbb{R}^{\mathbb{Z}}$  equipped with the  $T_P$ -invariant measure  $\mu_P$  defined by

$$\mu_P([A_1, A_2, \dots, A_n]_k) = \int_{\mathbb{R}} \tau(A_1, A_2, \dots, A_n) dm$$

where, for  $A_1, A_2, \dots, A_n \in \beta$ ,

and  $\tau = \tau_P$  is defined by

$$\tau(A_1) := 1_{A_1}, \quad \tau(A_1, A_2, \dots, A_n) := 1_{A_1} P\tau(A_2, \dots, A_n).$$

Note that the "consistency" and  $T_P$ -invariance of  $\mu_P$  follow from  $P1 = 1$  and  $\int_{\mathbb{R}} Pgd\bar{m} = \int_{\mathbb{R}} gdm$ .

Since  $f_\alpha$  has a strictly positive, continuous density, it follows that

$$\exists q \in (0, 1) \exists Pg \geq q 1_I \int_I gdm \forall g \in L_+^1(\mathbb{R}).$$

As in [AN], let  $X = \mathbb{R} \times \{0, 1\}$ ,  $\bar{m} = m \times (1 - q, q)$ , and define  $R: L^1(X) \rightarrow L^1(X)$  by

$$Rg := 1_{\mathbb{R} \times \{0, 1\}}(PEg) \circ \psi + 1_{I \times \{0\}} \frac{1}{1 - q} \left( (PEg) \circ \psi - q \int_I Egdm \right) + 1_{I \times \{1\}} \int_I Egdm$$

where  $E: L^1(X) \rightarrow L^1(\mathbb{R})$  is defined by  $Eg(x) := (1 - q)g(x, 0) + qg(x, 1)$ , and  $\psi: X \rightarrow \mathbb{R}$  is defined by  $\psi(x, \delta) = x$ . The choice of  $q$  ensures that  $R$  is a positive operator. It may be checked that  $R1 = 1$  and  $\int_{\mathbb{R}} Rgd\bar{m} = \int_{\mathbb{R}} gdm$ , and so  $T_R$ , the shift of  $R$  may be defined as above.

It follows from  $ERg = PEg$  that  $\pi: T_R \rightarrow T_P$ , where  $\pi: T^{\mathbb{Z}} = \mathbb{R}^{\mathbb{Z}} \times \{0, 1\}^{\mathbb{Z}} \rightarrow \mathbb{R}^{\mathbb{Z}}$  is the projection  $\pi(x, \varepsilon) = x$ .

We now show that  $w$  is a renewal sequence (for the chosen value of  $q$ ), and that  $T_R \rightarrow T_w$ . Set  $A = I \times \{1\}$ , and  $B = [A]_0$ . We claim that  $B$  is a recurrent event for  $T_R$  with renewal sequence  $w$ , whence lemma 1. It may be checked that

$$1_A Rg = 1_A \int_I Egdm,$$

whence

$$1_A R^{n+1} 1_A = 1_A \int_I ER^n 1_A d\bar{m} = q 1_A \int_I P^n 1_I d\bar{m}.$$

In particular, we have that

$$\mu_R(T_R^{-(n+1)} B | B) = \frac{1}{q} \int_X 1_A R^{n+1} 1_A d\bar{m} = w_{n+1}.$$

To show that  $B$  is a recurrent event for  $T_R$ , note that for  $0 = n_0 < n_1 < \dots < n_k$ ,

$$\bigcap_{j=0}^k T_R^{-n_j} B = [A, X^{m_1}, A, X^{m_2}, A, \dots, A, X^{m_k}, A]_0$$

where  $m_j = n_j - n_{j-1} - 1$ . Now,

$$\begin{aligned} \tau_R(A, X^{m_1}, A, X^{m_2}, A, \dots, A, X^{m_k}, A) &= \tau_R(A, X^{m_1}, A, X^{m_2}, A, \dots, A, X_{m_{k-1}}, A) w_{m_k} \\ &= \dots \prod_{j=1}^k w_{m_j} 1_A. \end{aligned}$$

This shows that  $B$  is a recurrent event for  $T_R$ .

*Proof of Lemma 2.* By lemma 5.2 of [AK], it is sufficient to show

$$w_n = \frac{1}{n^{1/\alpha}} \left( a + \frac{b}{n^{2/\alpha}} + \frac{c}{n^{4/\alpha}} + O\left(\frac{1}{n^{6/\alpha}}\right) \right)$$

as  $n \rightarrow \infty$  where  $a > 0, b, c \in \mathbb{R}$ .

To see this, let  $f_\alpha^{n*}$  denote the  $n$ -th convolution power of the probability  $f_\alpha$ , and note that

$$\begin{aligned} w_n &= \int_I \int_{\mathbb{R}} 1_I(x+y) df_\alpha^{n*}(y) dx \\ &= \int_{[-1,1]} (1-|y|) df_\alpha^{n*}(y) \\ &= \int_{\mathbb{R}} \int_{[-1,1]} (1-|y|) \cos ty dy e^{-n|t|^2} dt \\ &= 4 \int_0^\infty \phi(t) e^{-nt^2} dt \end{aligned}$$

where

$$\phi(t) = \frac{1 - \cos t}{t^2}$$

Changing variables,

$$w_n = \frac{4}{\alpha n^{1/\alpha}} \int_0^\infty \phi\left(\left(\frac{x}{n}\right)^{1/\alpha}\right) x^{(1/\alpha)-1} e^{-x} dx,$$

and we analyse the integral.

There exists  $r \in (0, 1)$  such that

$$\int_n^\infty \phi\left(\left(\frac{x}{n}\right)^{1/\alpha}\right) x^{(1/\alpha)-1} e^{-x} dx = O(r^n),$$

and

$$\int_n^\infty x^{(k/\alpha)-1} e^{-x} dx = O(r^n),$$

as  $n \rightarrow \infty$  for  $k \geq 1$ .

By Taylor's theorem

$$\phi(t) = a + bt^2 + ct^4 + \kappa(t)t^6$$

where  $\sup_{-1 \leq t \leq 1} |\kappa(t)| = M < \infty$ . It follows that

$$\begin{aligned} \int_0^n \phi\left(\left(\frac{x}{n}\right)^{1/\alpha}\right) x^{(1/\alpha)-1} e^{-x} dx &= \int_0^n \left( \sum_{k=0}^2 b_k \left(\frac{x}{n}\right)^{2k/\alpha} + \kappa\left(\frac{x}{n}\right) \left(\frac{x}{n}\right)^{6/\alpha} \right) x^{(1/\alpha)-1} e^{-x} dx \\ &= \sum_{k=0}^2 \frac{b'_k}{n^{2k/\alpha}} + O\left(\frac{1}{n^{6/\alpha}}\right) \end{aligned}$$

where  $\{b_k, b'_k : 0 \leq k \leq 2\}$  are constants and  $b'_0 > 0$ . This proves lemma 2.



## Acknowledgement

The author would like to thank the University of Paris for hospitality while preparing the paper.

## References

- [A1] Aaronson J, Rational ergodicity and a metric invariant for Markov shifts. *Israel J. Math.* **27** (1977) 93–123
- [A2] Aaronson J, The intrinsic normalising constants of transformations preserving infinite measures, *J. d'Analyse Math.* **49** (1987) 239–270
- [AK] Aaronson J and Keane M, Isomorphism of random walks, *Israel J. Math.* **87** (1994) 37–64
- [ALP] Aaronson J, Liggett T and Picco P, Equivalence of renewal sequences and isomorphism of and random walks, *Israel J. Math.* **87** (1994) 65–75
- [AN] Athreya K B and Ney P, A new approach to the limit theory of recurrent Markov chains, *TAMS* **248** (1978) 493–501
- [Ch] Chung K L, Markov Chains with stationary transition probabilities, Vol 104, Springer, Heidelberg, 1960
- [Fe] Feller W, An introduction to probability theory and its applications, volume I, (New York: John Wiley) (1968)
- [Kal] Kaluza T, Über die Koeffizienten reziproker Potenzreihen, *Math. Z.* **28** (1928), 161–170.
- [Kin] Kingman J F C, Regenerative Phenomena, (New York: John Wiley) (1972)



# $L^1(\mu, X)$ as a complemented subspace of its bidual

T S S R K RAO

Stat.-Math. Unit, Indian Statistical Institute, R V College Post, Bangalore 560059, India

MS received 5 August 1993; revised 28 September 1993

**Abstract.** We show that for a Banach space  $X$ , if the space of  $X$ -valued Bochner integrable functions is complemented in some dual space, then it is complemented in the space of  $X$ -valued countably additive,  $\mu$ -continuous vector measures.

**Keywords.** Bochner integrable functions; vector measures;  $L$ -ideals.

## 1. Introduction

Let  $(\Omega, \mathcal{A}, \mu)$  be a finite measure space and let  $X$  be a Banach space that is complemented in its bidual. In this note we consider the question “when is  $L^1(\mu, X)$  complemented in its bidual?”. It is well known that  $L^1(\mu)$  is complemented (by a norm one projection) in its bidual. Lindenstrauss [6] had observed that  $X$  is complemented in its bidual iff it is complemented in some dual space. Hence being complemented in its bidual is a property inherited by complemented subspaces. Using this observation, the author has proved in [7] that if  $X$  has the Radon Nikodym property w.r.t  $\mu$  and  $X$  is complemented in its bidual (which is clearly a necessary condition) then  $L^1(\mu, X)$  is complemented in its bidual. Looking at our argument Emmanuele [3] has recently pointed out that the stage where we use the R.N.P by requiring  $L^1(\mu, X) = cabv(\mu, X)$  (the space of countably additive  $X$ -valued measures on  $\mathcal{A}$  of bounded variation that are absolutely continuous w.r.t  $\mu$ ) can be replaced by the requirement  $L^1(\mu, X)$  is complemented in  $cabv(\mu, X)$  to yield the same conclusion. This becomes a significant remark in view of the recent result of F Freniche and L Rodriguez-Piazza, that for any Banach lattice not containing a copy of  $c_0$ ,  $L^1(\mu, X)$  is complemented in  $cabv(\mu, X)$  (see [3]).

In this note we point out that for an  $X$  that is complemented in its bidual,  $L^1(\mu, X)$  is complemented in its bidual iff it is complemented in  $cabv(\mu, X)$ . As a consequence we show that if  $L^1(\mu, X)$  is complemented in its bidual and  $M \subset X$  is a reflexive subspace then  $L^1(\mu, X|M)$  is complemented in its bidual. Our approach also gives easier proofs of some of the results from [3].

It should be noted that several authors [2, 8] have observed that if  $X$  has a copy of  $c_0$ , then  $L^1(\mu, X)$  is not complemented in its bidual.

**Main Result:** Let  $K$  denote the Stone space of  $L^\infty(\mu)$  and  $\wedge$  denote the Gelfand map. Let  $\hat{\mu}$  denote the measure defined on the Borel  $\sigma$ -field of  $K$  via the Gelfand map. It follows from the arguments given in [7] that the  $\wedge$  map can be extended to  $cabv(\mu, X)$  onto  $rcabv(\hat{\mu}, X)$  in such a way that  $L^1(\mu, X)$  gets mapped onto  $L^1(\hat{\mu}, X)$ . Therefore

**Theorem.** Let  $X$  be complemented in its bidual, then  $L^1(\mu, X)$  is complemented in its bidual iff it is complemented in  $cabv(\mu, X)$ .

*Proof.* From the argument given above it is clear that there is no loss of generality in assuming that  $\mu$  is a finite, category measure on the Borel  $\sigma$ -field of a hyperstonean space  $K$ .

Suppose  $L^1(\mu, X)$  is complemented in its bidual. Since  $L^1(\mu, X) = L^1(\mu) \hat{\otimes} X$  and  $L^1(\mu)^* = C(K)$  we have,

$$L^1(\mu, X)^{**} = \mathcal{L}(X, C(K))^*$$

(where  $\mathcal{L}(X, C(K))$  denotes the space of operators), see [1].

Let  $P: L^1(\mu, X)^{**} \rightarrow L^1(\mu, X)$  be a projection.

It is well known (see [5]) that

$$\mathcal{L}(X, C(K))^* = \mathcal{K}(X, C(K))^* \oplus \mathcal{K}(X, C(K))^\perp$$

(where  $\mathcal{K}(X, C(K))$  denotes the space of compact operators).

Since  $\mathcal{K}(X, C(K)) = C(K, X^*)$ , and by Singer's Theorem,  $C(K, X^*)^* = rcabv(X^{**})$ .

Therefore

$$L^1(\mu, X) \subset rcabv(\mu, X) \subset L^1(\mu, X)^{**}.$$

Hence  $P|rcabv(\mu, X) \rightarrow L^1(\mu, X)$  is the required projection.

The converse part of the proof proceeds in the same lines as the one in [7] where one now takes the composition with a projection from  $rcabv(\mu, X)$  into  $L^1(\mu, X)$  instead of the equality of these objects used in our earlier proof, as also pointed out during the proof of Theorem 6 in [3].

## COROLLARY 1

Suppose  $X$  is such that  $L^1(\mu, X)$  is complemented in its bidual. Let  $M \subset X$  be a reflexive subspace. Then  $L^1(\mu, X|M)$  is complemented in its bidual.

*Proof.* It is sufficient to show that  $L^1(\mu, X|M)$  is complemented in  $cabv(\mu, X|M)$  in the category measure set-up. Clearly  $X|M$  is complemented in its bidual.

It follows from Theorem 1 of [3] that what is required is to lift an  $F \in cabv(\mu, X|M)$  to an element of  $cabv(\mu, X)$ . Fix  $F \in cabv(\mu, X|M)$ . Since  $M \subset X^{**}$ ,  $F \in cabv(\mu, X^{**}|M)$ , by the proof of corollary 3 in [3], we get a lifting measure  $\tilde{F} \in cabv(\mu, X^{**})$ . Fix a  $A \in \mathcal{A}$  and let  $F(A) = \pi(x)$  where  $\pi$  is the quotient map and  $x \in X$ .

Since  $\pi(\tilde{F}(A)) = F(A) = \pi(x)$  we get that  $\tilde{F}(A) - x \in M$  and hence  $\tilde{F}(A) \in X$ .

Hence  $\tilde{F}$  is a lifting for  $F$ .

## COROLLARY 2

Let  $(\Omega, \mathcal{A}, \mu)$  and  $(\Omega', \mathcal{A}', \mu')$  be two finite measure spaces. Let  $\gamma$  denote the product measure on the product  $\sigma$ -field. Suppose  $X$  is complemented in its bidual. If  $L^1(\gamma, X)$  is complemented in  $cabv(\gamma, X)$  then  $L^1(\mu, L^1(\mu', X))$  is complemented in  $cabv(\mu, L^1(\mu', X))$ .

*Proof.* The hypothesis implies that  $L^1(\gamma, X)$  is complemented in its bidual. Therefore  $L^1(\mu', X)$  is complemented in its bidual. Let  $Y = L^1(\mu', X)$ . Since  $L^1(\mu, Y) = L^1(\gamma, X)$ , this space is complemented in its bidual and hence by the Theorem,  $L^1(\mu, Y)$  is complemented in  $cabv(\mu, Y)$ .

*Remark.* This provides a simple proof of Theorem 5 of [3] and our Corollary 1 improves on Corollary 4 of [3]. This also shows that one need not invoke the results of Freniche and Rodriguez-Piazza in the proof of Corollary 2 in [3].

Part of our motivation for looking at questions of this nature came from our interest in  $L$ -structure of Banach spaces. A Banach space  $X$  is said to be an  $L$ -ideal in its bidual, if there is an onto projection  $P: X^{**} \rightarrow X$  ( $X$  is canonically embedded in  $X^{**}$ ) such that  $\|P(\wedge)\| + \|\wedge - P(\wedge)\| = \|\wedge\|$ ,  $\forall \wedge \in X^{**}$ . For any positive measure  $\mu$ ,  $L^1(\mu)$  is such a space and an interesting open question in this area is whether  $L^1(\mu, X)$  is an  $L$ -ideal in its bidual whenever  $X$  is? We refer to the monograph [4] for properties of these spaces and for the authorship of some of the results that we will be quoting.

The result we are going to present below and some of the preceding results indicate the possibility that  $L^1(\mu, X)$  is a constrained subspace of its bidual, whenever  $X$  is an  $L$ -ideal in its bidual.

We need the following results from Chapter 4 of [4] due to D Li.

- 1) If  $X$  and  $Y$  are  $L$ -ideals in their biduals and  $Y \subset X$  (isometrically) then  $X|Y$  is an  $L$ -ideal in its bidual.
- 2) Let  $L$  be a Banach space such that  $L^*$  is injective and  $X, Y$  as in 1), then every operator  $T: L \rightarrow X|Y$  factors through the quotient map  $\pi: X \rightarrow X|Y$ .

## PROPOSITION

*Let  $X$  and  $Y$  be  $L$ -ideals in their biduals with  $Y \subset X$ . Suppose  $Y$  has the RNP and  $L^1(\mu, X)$  is constrained in its bidual, then  $L^1(\mu, X|Y)$  is constrained in its bidual.*

*Proof.* Note that because of 1),  $X|Y$  is an  $L$ -ideal in its bidual. It follows from the arguments given in [7] that w.l.o.g, we can assume that  $\mu$  is a finite measure.

It now follows from our Theorem that it is enough to show that  $L^1(\mu, X|Y)$  is constrained in  $cabv(\mu, X|Y)$ . In view of the results from [3], what is required to show is that any  $F \in cabv(\mu, X|Y)$  can be 'lifted' to a  $\tilde{F} \in cabv(\mu, X)$ . Let  $F \in cabv(\mu, X|Y)$  and define a bounded linear operator

$T: L^1(|F|) \rightarrow X|Y$  by the formula  $T([\chi_E]) = F(E)$  for any measurable set  $E$  where  $|F|$  is the variation of  $F$ .

Since  $L^\infty(|F|)$  is clearly an injective Banach space, it follows from the result 2) quoted above that  $\exists \tilde{T}: L^1(|F|) \rightarrow X$  such that

$$\pi \circ \tilde{T} = T$$

Now define  $\tilde{F}: \mathcal{A} \rightarrow X$  by  $\tilde{F}(E) = \tilde{T}([\chi_E])$ . From standard vector measure theory we know that  $\tilde{F} \in cabv(\mu, X)$  and

$$\begin{aligned} \pi(\tilde{F}(E)) &= \pi(\tilde{T}[\chi_E]) \\ &= T([\chi_E]) = F(E). \end{aligned}$$

## Acknowledgement

Thanks are due to Professor G Emmanuele for sending a copy of [3]. In the subsequent e-mail correspondence the author had with him, it was realized that he now also has a version of author's Theorem.

## References

- [1] Diestel J and Uhl J J, Vector measures, Math. Surveys. #15, *Am. Math. Soc. Providence* (1977)
- [2] Emmanuele G, On complemented copies of  $c_0$  in  $L_X^p$ , *Proc. Am. Math. Soc.* **104** (1988) 785–786
- [3] Emmanuele G, On the complementability of spaces of Bochner integrable functions in spaces of vector measures (preprint), May 1993
- [4] Harmand P, Werner D and Werner W, *M-ideal in Banach spaces and Banach algebras*, Springer LNH # 1547, Berlin 1993
- [5] Johnson J, Remarks on Banach spaces of compact operators, *J. Funct. Anal.* **32** (1979) 304–311
- [6] Lindenstrauss J, On a certain subspace of  $l_1$ , *Bull. De L'academie Polonaise Sci.* **9** (1964) 539–542
- [7] Rao T S S R K, A note on the  $R_{n,k}$  property for  $L^1(\mu, E)$ , *Can. Math. Bull.* **32** (1989) 74–77
- [8] Rao T S S R K, Intersection property of balls in tensor products of some Banach spaces-II, *Indian. J. Pure Appl. Math.* **21** (1990)

## Stresses in an elastic plate lying over a base due to strip-loading

RAJ KUMAR SHARMA and NAT RAM GARG

Department of Mathematics, Maharshi Dayanand University, Rohtak 124 001, India

MS received 27 January 1992; revised 25 October 1993

**Abstract.** The closed-form analytic expressions for the stresses at any point of an elastic plate coupling in different ways to a base as a result of a two-dimensional shear strip-loading are obtained. The contact between the horizontal layer and the base is either smooth-rigid or rough-rigid or welded. The variations of the shear stresses with the horizontal distance have been studied numerically. It is found that the effect of different boundary conditions on the stress field is significant and the stresses for an elastic layer lying over an elastic half-space differ considerably from those of an entire homogeneous elastic half-space.

**Keywords.** Elastic plate; rigid base; smooth base; shear stresses; shear surface loads.

### 1. Introduction

The solution of the problem of the deformation of a horizontally layered elastic material under the action of surface loads is finding wide applications in engineering, geophysics and soil mechanics. The deformation of a multilayered elastic half-space due to two-dimensional and three-dimensional surface loadings has been studied by many researchers, (Kuo [6], Singh [9], Pan [7], Chaudhuri and Bhowal [1], Garg and Sharma [2, 3] and others). Recently, Garg and Sharma [4] discussed the deformation of an elastic layer coupling in different ways to a base due to a very long vertical strike-slip fault situated in the layer.

In the present paper, the closed-form expressions for the stresses in an horizontal plate of infinite lateral extent lying over a base due to strip-loading are obtained. In geophysics, the elastic plate represents the crust of the earth. The interface between the plate and the base may be either welded or rigid. The rigid interface is either smooth-rigid or rough-rigid. The deformation of the plate corresponding to each type of the interface is obtained. The deformation of an homogeneous entire half-space due to strip-loading can be obtained from our results as a particular case. Similarly, the deformation of an elastic layer lying over an elastic half-space can be recovered from our results. Finally, the variation of stresses is studied numerically.

### 2. Basic equations

the displacement components are of the type

$$u = u(y, z), \quad v = w \equiv 0 \quad (1)$$

and  $u(y, z)$  satisfies the equilibrium equation

$$\frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0 \quad (2)$$

for zero body forces. The non-zero strains and stresses are

$$\begin{aligned} e_{12} &= \frac{1}{2} \frac{\partial u}{\partial y}, & e_{13} &= \frac{1}{2} \frac{\partial u}{\partial z} \\ \tau_{12} &= \mu \frac{\partial u}{\partial y}, & \tau_{13} &= \mu \frac{\partial u}{\partial z} \end{aligned} \quad (3)$$

$\mu$  being the rigidity of the medium.

### 3. Formulation of the problem

We consider a horizontal elastic plate of thickness  $H$  and rigidity  $\mu$  lying over a base. The origin of the cartesian coordinate system  $(x, y, z)$  is taken at the upper boundary of the plate and the  $z$ -axis is drawn into the medium. The plate occupies the region  $0 \leq z \leq H$  and the region  $z > H$  is the base over which the plate is lying (figure 1).

We assume that a shear-load  $R$  per unit area is acting over the strip  $|y| \leq a$  of the surface  $z = 0$  in the positive  $x$ -direction. The boundary condition at the surface  $z = 0$  is

$$\tau_{13} = \begin{cases} -R & |y| \leq a \\ 0 & |y| > a \end{cases} \quad (4)$$

The interface  $z = H$  between the plate and the base may be either smooth-rigid or rough-rigid or welded.

When the interface  $z = H$  is of the smooth-rigid type, the boundary condition at

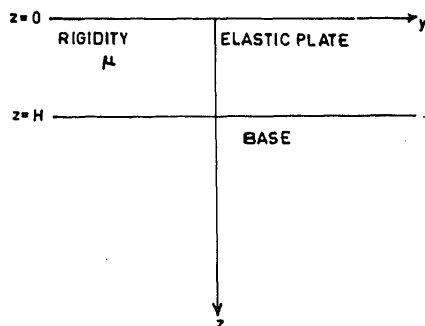


Figure 1. Section of the model by the plane  $x = 0$ .



$z = H$  is (Small and Booker [10])

$$\tau_{13}(z = H) = 0 \quad (5)$$

whereas for the rough-rigid interface, the boundary condition is (Small and Booker [10])

$$u(z = H) = 0. \quad (6)$$

When the interface  $z = H$  is welded, the continuity of the displacement and shear stress  $\tau_{13}$  implies (Singh [8])

$$\begin{aligned} u(z = H -) &= u(z = H +) \\ \tau_{13}(z = H -) &= \tau_{13}(z = H +). \end{aligned} \quad (7)$$

We shall find the deformation field at any point of the plate corresponding to each type of the boundary condition at the interface due to shear strip-loading.

#### 4. Solution of the problem

For the solution of (2), we make use of Fourier transform. The Fourier transform of  $X(y, z)$  is defined as

$$\bar{X}(k, z) = \int_{-\infty}^{\infty} X(y, z) \exp(iky) dy \quad (8)$$

so that

$$X(y, z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \bar{X}(k, z) \exp(-iky) dk. \quad (9)$$

Taking the Fourier transform of (2), we get

$$\left( \frac{d^2}{dz^2} - k^2 \right) \bar{u} = 0. \quad (10)$$

Solution of (10) is

$$\bar{u} = A \exp(-|k|z) + B \exp(|k|z) \quad (11)$$

where  $A$  and  $B$  may be functions of  $k$ . Then

$$u = \frac{1}{2\pi} \int_{-\infty}^{\infty} [A \exp(-|k|z) + B \exp(|k|z)] \exp(-iky) dk. \quad (12)$$

The shear stresses are

$$\tau_{12} = \frac{-i\mu}{2\pi} \int_{-\infty}^{\infty} [A \exp(-|k|z) + B \exp(|k|z)] \exp(-iky) k dk \quad (13)$$

$$\tau_{13} = \frac{\mu}{2\pi} \int_{-\infty}^{\infty} [-A \exp(-|k|z) + B \exp(|k|z)] \exp(-iky) |k| dk \quad (14)$$

Equations (4), (8) and (9) express the shear stress  $\tau_{13}$  in the integral form at the upper boundary  $z = 0$  of the plate. We find

$$\tau_{13} = \frac{-R}{\pi} \int_{-\infty}^{\infty} \frac{\sin ka}{k} \exp(-iky) dk. \quad (15)$$

Comparing (14) and (15), we obtain

$$A - B = \frac{2R}{\mu} \left( \frac{\sin ka}{k|k|} \right). \quad (16)$$

#### 4.1 Smooth-rigid boundary at $z = H$

When the interface  $z = H$  is smooth-rigid, the boundary condition (5) and eq. (14) yield

$$A \exp(-|k|H) - B \exp(|k|H) = 0. \quad (17)$$

Solving (16)–(17), we have

$$A = \frac{2R \sin ka}{\mu k|k|} \left[ \frac{1}{1 - \exp(-2|k|H)} \right], \quad B = \frac{2R \sin ka}{\mu k|k|} \left[ \frac{\exp(-2|k|H)}{1 - \exp(-2|k|H)} \right]. \quad (18)$$

The displacement and stresses for a smooth-rigid interface can be obtained directly from (12)–(14) and (18). Using the power series expansion for  $[1 - \exp(-2|k|H)]^{-1}$ , we obtain

$$u = \frac{R}{\pi\mu} \int_{-\infty}^{\infty} \sin ka \left[ \exp(-|k|z) + \sum_{n=1}^{\infty} \exp\{-|k|(2nH + z)\} \right. \\ \left. + \sum_{n=1}^{\infty} \exp\{-|k|(2nH - z)\} \right] \frac{\exp(-iky)}{k|k|} dk. \quad (19)$$

Using the Appendix, the closed-form expressions for the stresses  $\tau_{12}$  and  $\tau_{13}$  are obtained as follows:

$$\tau_{12} = \frac{-R}{2\pi} \left[ \log \left\{ \frac{(a+y)^2 + z^2}{(a-y)^2 + z^2} + \sum_{n=1}^{\infty} \left\{ \log \frac{(a+y)^2 + (2nH+z)^2}{(a-y)^2 + (2nH+z)^2} \right. \right. \right. \\ \left. \left. \left. + \log \frac{(a+y)^2 + (2nH-z)^2}{(a-y)^2 + (2nH-z)^2} \right\} \right] \right]. \quad (20)$$

$$\tau_{13} = \frac{-R}{\pi} \left[ \tan^{-1} \left( \frac{2az}{y^2 + z^2 - a^2} \right) + \sum_{n=1}^{\infty} \left\{ \tan^{-1} \left( \frac{2a(2nH+z)}{y^2 + (2nH+z)^2 - a^2} \right) \right. \right. \\ \left. \left. - \tan^{-1} \left( \frac{2a(2nH-z)}{y^2 + (2nH-z)^2 - a^2} \right) \right\} \right]. \quad (21)$$

$$A \exp(-|k|H) + B \exp(|k|H) = 0. \quad (22)$$

From (16) and (22), the values of  $A$  and  $B$  are found to be

$$A = \frac{2R \sin ka}{\mu k |k|} \left[ \frac{1}{1 + \exp(-2|k|H)} \right], \quad B = \frac{-2R \sin ka}{\mu k |k|} \left[ \frac{\exp(-2|k|H)}{1 + \exp(-2|k|H)} \right]. \quad (23)$$

The integral expression for the displacement, when the interface is rough-rigid, can be obtained from (12) and (23). We get

$$u = \frac{R}{\pi \mu} \int_{-\infty}^{\infty} \sin ka \left[ \exp(-|k|z) + \sum_{n=1}^{\infty} (-1)^n \exp\{-|k|(2nH+z)\} \right. \\ \left. + \sum_{n=1}^{\infty} (-1)^n \exp\{-|k|(2nH-z)\} \right] \frac{\exp(-iky)}{k|k|} dk. \quad (24)$$

The closed-form expressions for stresses are found to be

$$\tau_{12} = \frac{-R}{2\pi} \left[ \log \frac{(a+y)^2 + z^2}{(a-y)^2 + z^2} + \sum_{n=1}^{\infty} (-1)^n \left\{ \log \frac{(a+y)^2 + (2nH+z)^2}{(a-y)^2 + (2nH+z)^2} \right. \right. \\ \left. \left. + \log \frac{(a+y)^2 + (2nH-z)^2}{(a-y)^2 + (2nH-z)^2} \right\} \right] \quad (25)$$

$$\tau_{13} = \frac{-R}{\pi} \left[ \tan^{-1} \left( \frac{2az}{y^2 + z^2 - a^2} \right) + \sum_{n=1}^{\infty} (-1)^n \left\{ \tan^{-1} \left( \frac{2a(2nH+z)}{y^2 + (2nH+z)^2 - a^2} \right) \right. \right. \\ \left. \left. - \tan^{-1} \left( \frac{2a(2nH-z)}{y^2 + (2nH-z)^2 - a^2} \right) \right\} \right]. \quad (26)$$

#### 4.3 Welded interface

Let  $\mu_0$  be the rigidity of the half-space  $z > H$ . The displacement  $u$  in the region  $z \geq H$  is of the type

$$u = \frac{1}{2\pi} \int_{-\infty}^{\infty} A_0 \exp(-|k|z) \exp(-iky) dy \quad (27)$$

in which the coefficient  $A_0$  is to be determined from the boundary conditions. Then

$$\tau_{13} = -\frac{\mu_0}{2\pi} \int_{-\infty}^{\infty} A_0 \exp(-|k|z) \exp(-iky) |k| dk. \quad (28)$$

Equations (7), (12), (14), (27) and (28) yield the relations

$$A \exp(-|k|H) + B \exp(|k|H) = A_0 \exp(-|k|H) \quad (29)$$

$$\mu[-A \exp(-|k|H) + B \exp(|k|H)] = -\mu_0 A_0 \exp(-|k|H). \quad (30)$$

Solving (16), (29) and (30), we get

$$\begin{aligned} A_0 &= \frac{4R}{\mu} \left[ \frac{1}{1 - M \exp(-2|k|H)} \right] \left( \frac{S}{S+1} \right) \frac{\sin ka}{k|k|} \\ A &= \frac{2R}{\mu} \left[ \frac{1}{1 - M \exp(-2|k|H)} \right] \frac{\sin ka}{k|k|} \\ B &= \frac{2R}{\mu} \left[ \frac{M \exp(-2|k|H)}{1 - M \exp(-2|k|H)} \right] \frac{\sin ka}{k|k|} \end{aligned} \quad (31)$$

where

$$S = \mu/\mu_0, \quad M = (S-1)/(S+1) = (\mu - \mu_0)/(\mu + \mu_0). \quad (32)$$

Using (31), we obtain the deformation field for the welded interface as follows:

$$\begin{aligned} u &= \frac{R}{\pi\mu} \int_{-\infty}^{\infty} \left[ \exp(-|k|z) + \sum_{n=1}^{\infty} M^n \exp\{(-|k|(2nH+z))\} \right. \\ &\quad \left. + \sum_{n=1}^{\infty} M^n \exp\{-|k|(2nH-z)\} \right] \frac{\sin ka}{k|k|} \exp(-iky) dk \end{aligned} \quad (33)$$

$$\begin{aligned} \tau_{12} &= \frac{-R}{2\pi} \left[ \log \frac{(a+y)^2 + z^2}{(a-y)^2 + z^2} + \sum_{n=1}^{\infty} M^n \left\{ \log \frac{(a+y)^2 + (2nH+z)^2}{(a-y)^2 + (2nH+z)^2} \right. \right. \\ &\quad \left. \left. + \log \frac{(a+y)^2 + (2nH-z)^2}{(a-y)^2 + (2nH-z)^2} \right\} \right] \end{aligned} \quad (34)$$

$$\begin{aligned} \tau_{13} &= \frac{-R}{\pi} \left[ \tan^{-1} \left( \frac{2az}{y^2 + z^2 - a^2} \right) + \sum_{n=1}^{\infty} M^n \left\{ \tan^{-1} \left( \frac{2a(2nH+z)}{y^2 + (2nH+z)^2 - a^2} \right) \right. \right. \\ &\quad \left. \left. - \tan^{-1} \left( \frac{2a(2nH-z)}{y^2 + (2nH-z)^2 - a^2} \right) \right\} \right]. \end{aligned} \quad (35)$$

We observe that the deformation fields for the smooth-rigid and rough-rigid interfaces can also be obtained from the deformation field for the welded interface, respectively, on substituting  $M = 1$  and  $M = -1$ .

## 5. Numerical results

We study numerically the variation of the shear stresses  $\tau_{12}$  and  $\tau_{13}$  with the horizontal distance  $y$  for fixed values of other quantities  $a$ ,  $S$  and  $z$ . The width of the strip is assumed to be  $0.2H$ . It has been seen earlier that the values of  $S$  affect the shear stresses only when the interface between the elastic plate and the base is welded.

Figures 2 and 3 exhibit the variation of the stress  $\tau_{12}$ . Two values of the ratio  $S$  of rigidities are taken, namely,  $S = 0.25$  and  $S = 4.0$ .  $S = 0.25$  implies that the rigidity of the elastic plate is less than that of the base and  $S = 4.0$  means that the elastic plate is of high rigidity as compared to that of the base. Stresses at the points at

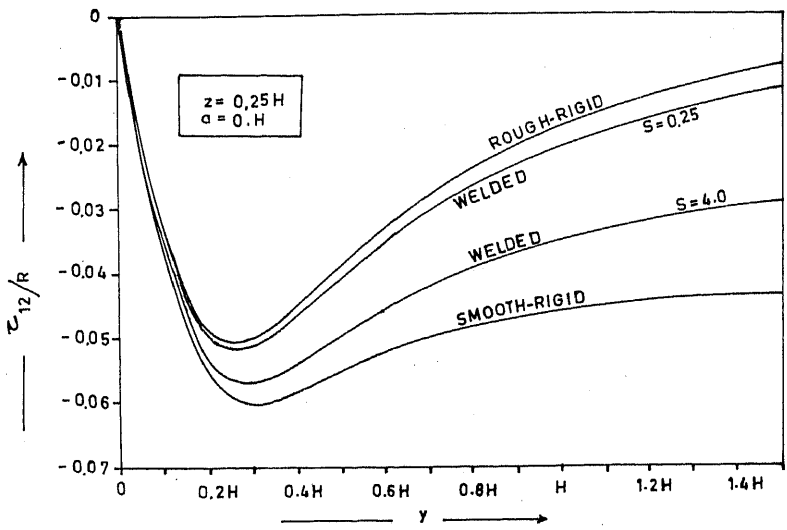


Figure 2. Variation of the dimensionless stress  $\tau_{12}/R$  with the horizontal distance  $y$  when  $a = 0.1H$  and  $z = 0.25H$ .

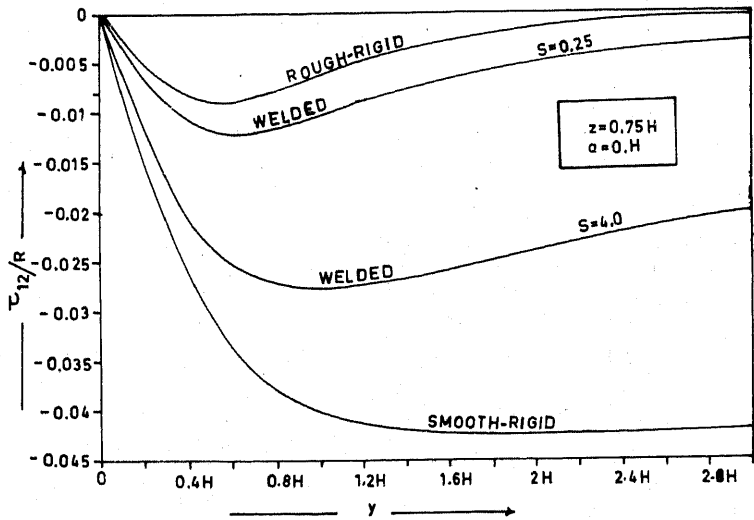


Figure 3. Variation of the dimensionless stress  $\tau_{12}/R$  with the horizontal distance  $y$  when  $a = 0.1H$  and  $z = 0.75H$ .

depths  $H/4$  and  $3H/4$  are calculated. It is observed that the stress  $\tau_{12}$  for the welded contact lies between the corresponding values of  $\tau_{12}$  for the rough-rigid and smooth-rigid contacts, for different values of  $S$ .

Figures 4 and 5 show the variation of the stress  $\tau_{13}$  with the distance  $y$ . In these figures, curves corresponding to different kinds of coupling of the elastic plate to a base have been drawn.

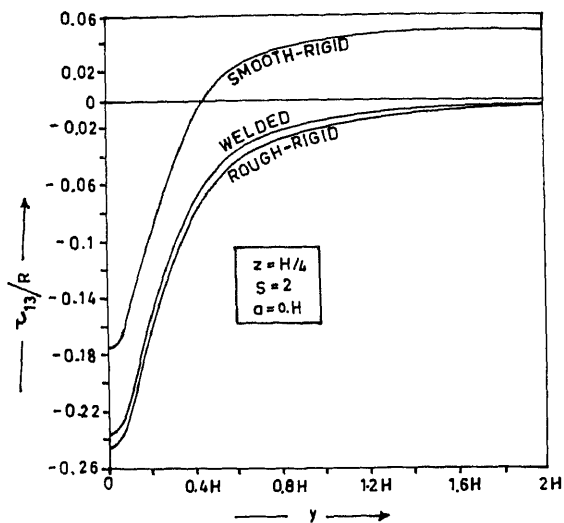


Figure 4. Variation of dimensionless stress  $\tau_{13}/R$  with  $y$  when  $a = 0.1H$ ,  $z = H/4$  and  $S = 2$ .

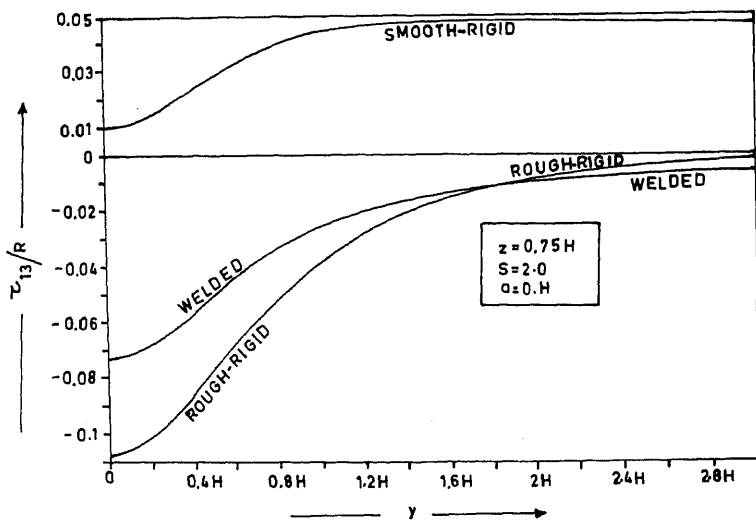


Figure 5. Variation of  $\tau_{13}/R$  with  $y$  when  $a = 0.1H$ ,  $z = 0.75H$  and  $S = 2$ .

### Acknowledgement

One of the authors (NRG) is thankful to the University Grants Commission, New Delhi, for providing financial support. The authors are also grateful to the referee for providing valuable suggestions for the improvement of the paper.

Appendix ( $\xi > 0$ )

$$(1) \int_{-\infty}^{\infty} \exp(-|k|\xi) \frac{\sin ka}{|k|} \exp(-iky) dk = \frac{-i}{2} \log \left[ \frac{(y+a)^2 + \xi^2}{(y-a)^2 + \xi^2} \right]$$

$$(2) \int_{-\infty}^{\infty} \exp(-|k|\xi) \frac{\sin ka}{k} \exp(-iky) dk = \tan^{-1} \left( \frac{2a\xi}{y^2 + \xi^2 - a^2} \right)$$

In integral (2), it is assumed that  $y^2 + \xi^2 \geq a^2$ . If, however,  $y^2 + \xi^2 < a^2$ , the quantity  $\pi$  is to be added on the right side [5].

## References

- [1] Chaudhuri P K and Bhowal S, Two-dimensional static response of a transversely isotropic multilayered non-homogeneous half-space to surface-loads, *Geophys. Res. Bull.* **27** (1989) 77-87
- [2] Garg N R and Sharma R K, Generation of displacements and stresses in a multilayered half-space due to strip-loading, *Bull. Indian Soc. Earthquake Technol.* **28** (1991) 1-26
- [3] Garg N R and Sharma R K, Displacement and stresses at any point of a transversely isotropic multilayered half-space due to strip-loadings, *Indian J. Pure Appl. Math.* **22** (1991) 859-877
- [4] Garg N R and Sharma R K, Deformation of an elastic layer coupling in different ways to a base due to a very long vertical strike-slip dislocation, *Proc. Indian Acad. Sci. (Earth Planet. Sci.)* **101** (1992) 255-268
- [5] Gradshteyn I S and Ryzhik I M, *Tables of Integrals, Series and Products* (New York: Academic Press) (1980)
- [6] Kuo J T, Static response of a multilayered medium under inclined surface loads, *J. Geophys. Res.* **74** (1969) 3195-3207
- [7] Pan E, Static response of a transversely isotropic and layered half-space to general surface loads, *Phys. Earth Planet. Inter.* **54** (1989) 353-363
- [8] Singh S J, Static deformation of a multilayered half-space by internal sources, *J. Geophys. Res.* **75** (1970) 3257-3263
- [9] Singh S J, Static deformation of a transversely isotropic multilayered half-space by surface loads, *Phys. Earth Planet. Inter.* **42** (1986) 263-73
- [10] Small J C and Booker J R, Finite layer analysis of layered elastic materials using a flexibility approach. Part-I-strip-loading, *Int. J. Num. Methods Eng.* **20** (1984) 1025-1037





# The Laplacian on algebraic threefolds with isolated singularities

VISHWAMBHAR PATI

Indian Statistical Institute, 8th Mile, Mysore Road, Bangalore 560 059, India

MS received 1 September 1993

**Abstract.** We give a complete description of the induced Fubini-Study metric (up to quasi-isometry) in a neighbourhood of an isolated complex projective threefold singularity, by using a sufficiently high resolution of singularities. This is then used to prove the self-adjointness of the corresponding Laplacian acting on square integrable functions, on the non-compact smooth locus of complex projective threefolds with isolated singularities.

**Keywords.** Laplacian; singularities; three folds; Fubini-study metric; self-adjointness.

## 1. Introduction

We consider a complex projective algebraic threefold  $X$  with isolated singularity set  $\Sigma$ . On  $X - \Sigma$ , which is non-compact, we consider the induced (incomplete) Fubini-Study metric  $g$ , and the corresponding Laplacian  $\bar{\Delta} = \bar{\partial}\bar{\partial}^*$  on  $L^2$ -functions with respect to this metric. The main theorem of the following is:

**Theorem 1.1.** *On  $X - \Sigma$  as above, the Laplacian  $\bar{\Delta} = \bar{\partial}\bar{\partial}^*$  is self-adjoint on its domain.*

In general, incomplete metrics on non-compact manifolds do not lead to self-adjoint Laplacians. In [2], [3] the self-adjointness of the Laplacian on  $i$ -forms has been established for admissible Riemannian pseudomanifolds, (covering algebraic curves). For singular algebraic surfaces, this was proved by Nagase [6] for the Laplacian acting on functions. This paper falls naturally into two parts. In the first (§2) we analyse the induced Fubini metric in a neighbourhood of the singular set and make it quasi-isometric to simpler models, using resolution of singularities, embedded point and curve blow-ups. In §3 we break up the proof of (1.1), which is essentially a norm estimate on  $\text{dom } \bar{\Delta}$ , to estimate on the corresponding simplified models (c.f. [6], Lemma 5.2). The paper [8], uses the results of this one to arrive at the trace estimate for the associated heat operator.

## 2. Reduction of the metric

This section constitutes an unpublished part of the author's Ph D thesis [7] and is reviewed here for completeness. The main idea, as in [5], is to obtain a decomposition of a  $b$ -neighbourhood of a singular point into product regions of the type  $(0, b) \times N_\alpha$ , where  $N_\alpha$  is a neighbourhood on the link of the singular point. These regions arise naturally from a covering of the exceptional divisor  $E = \pi^{-1}(\Sigma)$ , in a sufficiently high

resolution of singularities (to be described in the sequel), by polydisc neighbourhood. It is then shown that to obtain the basic estimate for self-adjointness (see [2], [3], [6]), it is enough to do it on each of these polydisc neighbourhoods.

## 2.1 Blowing up points and curves

In the context of three-folds, blowing up (a neighbourhood  $U$  of) the point

$$0 \in \{(u, v, w) \in \mathbb{C}^3\}$$

leads to a three-fold covered by three charts with coordinate systems  $(u_i, v_i, w_i)$  defined respectively by:

$$\begin{array}{lll} u = u_1 & u = u_2 v_2 & u = u_3 w_3 \\ v = u_1 v_1 & v = v_2 & v = v_3 w_3 \\ w = u_1 w_1 & w = v_2 w_2 & w = w_3 \end{array}$$

These are nothing but the defining equations of the resolution map  $\pi$ . The origin is replaced by an exceptional divisor  $\pi^{-1}(0) = \mathbb{CP}(2)$  defined by  $(u_1 = v_2 = w_3 = 0)$ . Hence if we have a monomial  $u^a v^b w^c$  it would transform, respectively, into three monomials  $u_1^{a+b+c} v_1^b w_1^c$ ,  $u_2^a v_2^{a+b+c} w_2^c$ ,  $u_3^a v_3^b w_3^{a+b+c}$  in the three new charts.

### DEFINITION 2.1.1

Let us define a list to be an  $N \times 3$  matrix of non-negative integers, such that some column contains no zero entry.

The reason for this technical condition is as follows. The primary list we shall be concerned with will be the exponents  $(a_i, b_i, c_i)_{i=1}^N$ , where the local coordinate function in  $\mathbb{C}^N$  (at whose origin our 3-fold singularity germ  $X$  is sitting)  $\{z_i\}_{i=1}^N$  are given by

$$z_i = u^{a_i} v^{b_i} w^{c_i} h_i(u, v, w),$$

where  $(u, v, w)$  is a local coordinate system centred at a point  $p = (0, 0, 0)$  of the singular divisor  $E$  (with normal crossings and no self-intersection) in a resolution of singularity of  $X$ . By convention (see § 2.2 below),  $E$  is defined in this local coordinate system by one or more of the coordinate hyperplanes  $(u = 0)$ ,  $(v = 0)$ ,  $(w = 0)$ , so that all the  $h_i$  vanish along one or more of these hyperplanes, which clearly implies that one of the columns  $(a_i)$ ,  $(b_i)$ ,  $(c_i)$  contains no zero entry.

All the lists  $(a'_i, b'_i, c'_i)_{i=1}^N$  we will be considering in this paper will dominate the primary list  $(a_i, b_i, c_i)_{i=1}^N$ , which is defined to mean that

$$a'_i \geq a_i, \quad b'_i \geq b_i, \quad c'_i \geq c_i$$

for all  $1 \leq i \leq N$ . So any list we consider will also have a column devoid of zero therefore.

### DEFINITION 2.1.2

We define a point blow-up, or type A operation of a list  $(a_i, b_i, c_i)_{i=1}^N$  to be the three new lists

$$(a_i + b_i + c_i, b_i, c_i)_{i=1}^N, (a_i, a_i + b_i + c_i, c_i)_{i=1}^N \text{ and } (a_i, b_i, a_i + b_i + c_i)_{i=1}^N.$$

We may also blow up closed curves lying on the three-fold. If a curve is given in a chart by the equations ( $v = w = 0$ ), then, blowing up this curve (whose local germ is the  $u$ -axis) would give two new charts, with coordinates  $(u_1, v_1, w_1)$  and  $(u_2, v_2, w_2)$  and the equations defining the resolution map  $\pi$  in the two new charts are, respectively,

$$\begin{array}{ll} u = u_1 & u = u_2, \\ v = v_1 w_1 & v = v_2, \\ w = w_1 & w = v_2 w_2. \end{array}$$

This would replace the  $u$ -curve (i.e. the  $u$ -axis)  $C$  by a  $C \times CP(1)$ .

### DEFINITION 2.1.3

A type B operation of the  $u$ -axis on the list  $(a_i, b_i, c_i)$  is defined to be the two new lists

$$(a_i, b_i + c_i, c_i), \text{ and } (a_i, b_i, b_i + c_i)$$

There are two points of caution, however. Firstly, if we blow up a curve, we must be sure to do the type B operation in *all* the charts covering it. Secondly, we must blow up only *closed* curves which *lie on the singular set*. So, if for example, the singular set is described by ( $u = 0$ ), we cannot blow up the  $u$ -curve. This is because such a curve would map down to a curve passing through the origin and meeting the non-singular part  $X - \{0\}$  of the three-fold. The inverse map  $\pi^{-1}$  (where  $\pi$  is the resolution map) will not be well defined on such points, and thus not a biholomorphism on  $X - \{0\}$ . Hence if for a list  $(a_i, b_i, c_i)$  we have  $b_i = c_i = 0$  for some  $i \in \{1, \dots, N\}$ , then the same holds for the primary list defined above in 2.1.1. This means that the  $u$ -axis ( $v = w = 0$ ) is not part of the singular divisor  $E$  (which is therefore defined locally by  $u = 0$ ). Thus we cannot blow up the  $u$ -axis as above. Thus

### DEFINITION 2.1.4

A permissible type B operation on a list is defined to be one which does not blow up the  $i$ th axis if, on deleting the  $i$ th column, the remaining two columns have a zero entry on the same row.

This restriction does not depend on local charts, though, since the condition that a curve lie inside or out of the singular divisor is independent of charts.

**Lemma 2.1.5.** *Given a list  $(a_i, b_i, c_i)_{i=1}^N$ , repeated permissible operations of type A and B may be performed, so that in all of the final lists, say  $(a_i, b_i, c_i)$ ,  $a_i > a_j$  implies that  $b_i \geq b_j$  and  $c_i \geq c_j$  for all pairs  $1 \leq i, j \leq N$ .*

*Proof.* For brevity, let us say a list is *arranged* if it satisfies the conclusion of the lemma. As in the case of two dimensions (c.f. [5]), it suffices to consider an array

$$a_1 \quad b_1 \quad c_1$$

so it is enough to tackle a pair of rows at a time). It is thus enough to rule out the situation  $a_1 < a_2, b_1 \geq b_2, c_1 \geq c_2$ , which is the only bad situation possible, upto a permutation of the  $a$ 's,  $b$ 's and  $c$ 's. We may further assume that  $b_1 > b_2, c_1 > c_2$ , because if two of them are equalities, there is nothing to prove. On the other hand if only one is an equality, say  $c_1 = c_2$ , and the other two are strict inequalities,  $a_2$  and  $b_1$  have to be  $> 0$ , so there is no restriction to a type B operation on the first and second columns and have them arranged as in Lemma 2.1 in [5]. So without loss of generality, we may assume  $a_1 < a_2, b_1 > b_2, c_1 > c_2$ . The idea is to induct on the negative integer  $a_1 - a_2$  until it becomes positive.

#### DEFINITION 2.1.6

An array where the order of the  $a_i, b_i, c_i$  is preserved, but this integer has (strictly) increased will be called an inductive improvement on the old array.

Let us assume (proof later)

*Condition 1.*  $a_2 - a_1 \geq b_1 - b_2; a_2 - a_1 \geq c_1 - c_2$

If any equality holds in Condition 1, say  $a_2 - a_1 = b_1 - b_2$  so that  $a_1 + b_1 = a_2 + b_2$ , then a type-B operation gives the two arrays, the first being

$$\begin{array}{ccc} a_1 + b_1 & b_1 & c_1 \\ a_2 + b_2 & b_2 & c_2 \end{array}$$

which is arranged, and we are through. The second array is

$$\begin{array}{ccc} a_1 & a_1 + b_1 & c_1 \\ a_2 & a_2 + b_2 & c_2 \end{array}$$

in which the middle column entries are equal, and hence the column may be ignored, and type B operations may be applied to the other two columns (as in Lemma 2.1, [5]) to get the result. These are permissible, (cf. 2.1.4 for the definition) since  $a_2$  and  $b_1 > 0$  implies that the middle column entries are both non-zero, and by our earlier assumption, both  $a_2$  and  $c_1$  are strictly positive. So, we may as well assume the stronger

*Condition 1'.*  $a_2 - a_1 > b_1 - b_2, a_2 - a_1 > c_1 - c_2$

Assuming this, and making a type A operation gives the three arrays

$$\begin{array}{ccc} a_1 + b_1 + c_1 & b_1 & c_1 \\ a_2 + b_2 + c_2 & b_2 & c_2 \end{array} \quad (1)$$

$$\begin{array}{ccc} a_1 & a_1 + b_1 + c_1 & c_1 \\ a_2 & a_2 + b_2 + c_2 & c_2 \end{array} \quad (2)$$

$$\begin{array}{ccc} a_1 & b_1 & a_1 + b_1 + c_1 \\ a_2 & b_2 & a_2 + b_2 + c_2 \end{array} \quad (3)$$

Now, in (1),

$$a'_1 - a'_2 = (a_1 - a_2) + (b_1 - b_2) + (c_1 - c_2) \geq (a_1 - a_2) + 1$$

since  $b_1 > b_2$ . The  $b$ 's and  $c$ 's are unchanged.

In (2),  $a_1 < a_2$  but

$$b'_1 - b'_2 = b_1 - b_2 + [(a_1 - a_2) + (c_1 - c_2)] < b_1 - b_2$$

by condition 1'. Also  $c_1 \geq c_2$  remain unchanged.

In (3), similar to (2),

$$a_1 < a_2, c'_1 - c'_2 < c_1 - c_2, \text{ and } b_1 \geq b_2$$

So, to sum up, we have an inductive improvement in (1), whereas in (2) and (3) condition 1' is preserved, with the same (negative)  $a_1 - a_2$ . Hence if we repeat type A operations, we have arrays

$$\begin{array}{ccc} a_1^{(i)} & b_1^{(i)} & c_1^{(i)} \\ a_2^{(i)} & b_2^{(i)} & c_2^{(i)} \end{array} \quad (4)$$

with an inductive improvement on the original array, or if  $a_1^{(i)} = a_1, a_2^{(i)} = a_2$ , then by the above, one of the two differences  $b_1^{(i)} - b_2^{(i)}, c_1^{(i)} - c_2^{(i)}$  strictly decreases (i.e. the cases (2), (3) respectively, above) at the  $i$ th step. So, after a finite number of steps, there are only the following three possibilities

$$a_1 < a_2, b_1^{(i)} \leq b_2^{(i)}, c_1^{(i)} \leq c_2^{(i)}, \quad (5)$$

$$a_1 < a_2, b_1^{(i)} \leq b_2^{(i)}, c_1^{(i)} \geq c_2^{(i)}, \quad (6)$$

$$a_1 < a_2, b_1^{(i)} \geq b_2^{(i)}, c_1^{(i)} \leq c_2^{(i)}. \quad (7)$$

In case of (5), the array is arranged, and we are done. In case of (6) let us redefine by changing the subscripts, i.e.

$$\tilde{a}_1 = a_2, \tilde{b}_1 = b_2^{(i)}, \tilde{c}_1 = c_2^{(i)}$$

and

$$\tilde{a}_2 = a_1, \tilde{b}_2 = b_1^{(i)}, \tilde{c}_2 = c_1^{(i)}$$

Thus (6) above becomes  $\tilde{b}_1 \geq \tilde{b}_2, \tilde{a}_1 \geq \tilde{a}_2, \tilde{c}_1 \leq \tilde{c}_2$ . Now note that

$$\tilde{c}_2 - \tilde{c}_1 = c_1^{(i)} - c_2^{(i)} \leq c_1 - c_2$$

because of the discussion about the case (3) above and

$$\tilde{c}_1 - \tilde{c}_2 = c_2^{(i)} - c_1^{(i)} > a_1^{(i)} - a_2^{(i)} = a_1 - a_2 \quad (8)$$

where the middle inequality comes because condition 1' persists for the array (4). So if we interchange roles of  $a$ 's and  $c$ 's, (8) above implies that we have an inductive improvement. The argument for the case (7) is the same as the above one for the (6) after exchanging the roles of  $b$ 's and  $c$ 's.

It therefore remains only to obtain the condition 1. As explained at the outset of the proof, we may take  $a_1 < a_2, b_1 > b_2, c_1 > c_2$ . To get condition 1, we perform type B operations. For example, doing this on the first two columns (which is clearly permissible by the assumption above on the  $a$ 's,  $b$ 's and  $c$ 's) leads to the arrays

$$\begin{array}{ccc} a_1 + b_1 & b_1 & c_1 \\ a_2 + b_2 & b_2 & c_2 \end{array}$$

In the first array above,

$$a'_1 - a'_2 = (a_1 - a_2) + (b_1 - b_2) > a_1 - a_2$$

where the last inequality is by the assumption  $b_1 > b_2$ . So we have an inductive improvement. In the second array above, we have

$$b'_1 - b'_2 = (b_1 - b_2) + (a_1 - a_2) < b_1 - b_2$$

by the assumption  $a_1 < a_2$ . Also  $a_i$  remain untouched. So eventually we obtain  $a_2 - a_1 > b_1^{(k)} - b_2^{(k)}$  which is the first inequality in condition 1. Similarly using type B operations on the first and third columns we obtain  $a_2 - a_1 > c_1 - c_2$ , and hence condition 1. Observe that in these inductive steps towards condition 1, the negative integer  $a_1 - a_2$  either improves or stays the same. Hence the lemma.  $\square$

## 2.2 Standard form of local parametrization

It is known from Hironaka [4] that an isolated singularity of a three-fold can be resolved by repeated embedded blow-ups along submanifolds. We let  $X$  denote the germ of the isolated 3-fold singularity at the origin of  $\mathbb{C}^N$  where  $N \geq 4$ , and  $\pi: \tilde{X} \rightarrow X$  be the resolution map thus guaranteed. It can also be arranged that if  $0 \in X$  is the singularity, then the exceptional divisor  $E = \pi^{-1}(0)$  has normal crossings and no self intersection. Further, we note

$$\pi: (\tilde{X} - E) \rightarrow (X - \{0\})$$

is a biholomorphic map.

We let  $\tilde{D}_i = \pi^* D_i$ , respectively  $\tilde{E}_{ijk} = \pi^* E_{ijk}$  (where  $i, j, k$  runs over all subsets of  $\{1, 2, \dots, N\}$  of cardinality 3) denote the *strict transforms*\* of the surfaces

$$D_i = \{z_i = 0\} \cap (X - \{0\})$$

and

$$E_{ijk} = \{dz_i \wedge dz_j \wedge dz_k = 0\} \cap (X - \{0\})$$

under the resolution map  $\pi$ . By a good choice of the coordinate functions  $\{z_i\}_{i=1}^N$ , on  $\mathbb{C}^N$  we have ensured that all the intersections among the various  $\tilde{D}_i$  and  $\tilde{E}_{ijk}$  away from  $E$  are normal crossings (see the first Proposition 4.1.1 in the Appendix). Since  $\pi$  is biholomorphic outside  $E$ , these intersections continue to be normal crossings under further blow-ups of points and curves on  $E$ . Now, by further blow-ups of points and curves on  $E$ , one also ensures that these  $\tilde{D}_i$  and  $\tilde{E}_{ijk}$  are also smooth, and cross normally with  $E$ .

Thus, from the above discussion, if  $p \in E = \pi^{-1}(0)$  is a point on the exceptional divisor, at most three components of  $E$  will meet at  $p$ . Thus such a point  $p$  can be of three types:

*Simple points.* At such a point  $p$ , exactly one irreducible component of the singular divisor  $E$  passes through  $p$ . We can choose a coordinate system  $(u, v, w)$  based at  $p = (0, 0, 0)$  such that locally  $u$  defines (this component of)  $E$  (set-theoretically) as  $(u = 0)$ . Multiplying  $u$  by local units will not change this local description of  $E$  as  $(u = 0)$ .

\*The strict transform of a subvariety  $Y$  of  $X - \{0\}$  is the closure of  $\pi^{-1}(Y)$  in  $\tilde{X}$ .

*Double points.* At such a point  $p$ , two components of  $E$  pass through  $p$ . We can choose local coordinates  $(u, v, w)$  based at  $p = (0, 0, 0)$  so that  $E$  is defined locally (set-theoretically) by  $(uv = 0)$ . Again, multiplying  $u, v$  does by local units not alter this local description.

*Triple points.* At such a point  $p$ , three components of  $E$  pass through  $p$ . Again, we can choose a coordinate system  $(u, v, w)$  based at  $p = (0, 0, 0)$  such that  $E$  is locally defined set-theoretically by  $(uvw = 0)$ . Multiplying these coordinates by local units does not change this local description of  $E$ .

We will have to analyse these three kinds of points on  $E$  separately. At this point it is useful to make the definition:

## DEFINITION 2.2.1

In the sequel, we shall be individually considering, case by case, charts of various types, centred at simple, double or triple points. These points will themselves fall into substrata (of the set of simple, double or triple points, respectively). Charts centred at such points will therefore also fall into the corresponding classification (i.e., all the cases and subcases treated below). When we perform a (permissible) type A or B operation on such a chart of a particular classification, the resulting charts of the same classification will be called relevant to that case/subcase. Clearly, in proving any of the succeeding propositions 2.2.2, 2.2.8, or 2.2.11, for charts of particular case or subcase, it is enough to prove them for *all* the relevant charts to that case or subcase.

**2.21 The analysis of simple-point charts** In this subsection we look at simple points. So let  $p$  be a simple point on the exceptional divisor  $E = \pi^{-1}(0)$ . The main proposition is the following:

## PROPOSITION 2.2.2

*After sufficiently many permissible type A (point blow-ups) and type B (closed curve blow-ups) operations, for  $p$  a point in  $S$ , there exists a coordinate system  $(u, v, w)$  centered at  $p$  in a polydisc neighborhood  $U$  centred at  $p = (0, 0, 0)$  such that  $U \cap E = (u = 0)$  and after permuting the coordinate functions and rescaling them, we have*

$$\begin{aligned} z_1 &= \zeta_1 \\ z_2 &= f'_2(\zeta_1) + \zeta_2 \\ z_3 &= f'_3(\zeta_1, \zeta_2) + \zeta_3 \\ z_i &= f'_i(\zeta_1, \zeta_2, \zeta_3) \quad (4 \leq i \leq N) \end{aligned} \tag{9}$$

where  $a_3 \geq a_2 \geq a_1$  are positive integers  $\geq 1$ ,  $\zeta_1 = u^{a_1}$ ,  $\zeta_2 = u^{a_2}v$ ,  $\zeta_3 = u^{a_3}w$ .  $f'_i$  are infinite series in fractional powers of  $\zeta_i$ . Substituting these expressions for  $\zeta_i$  in  $f'_i$ , we get holomorphic power series  $f_i$  in  $(u, v, w)$  such that

- (i) every monomial in  $f_i$  for  $(2 \leq i \leq N)$  is divisible by  $\zeta_1 = u^{a_1}$ ,
- (ii) every monomial in  $f_i$  for  $(3 \leq i \leq N)$  containing  $v$  is divisible by  $\zeta_2 = u^{a_2}v$ , and finally,
- (iii) every monomial in  $f_i$  for  $(4 \leq i \leq N)$  containing  $w$  is divisible by  $\zeta_3 = u^{a_3}w$ .

$S^2 = S - C$ , where

$$C = \left( \bigcup_i \tilde{D}_i \right) \cup \left( \bigcup_{ijk} \tilde{E}_{ijk} \right)$$

$S^1 = S \cap (C - A)$ , where

$$A = \left( \bigcup_{l,m} (\tilde{D}_l \cap \tilde{D}_m) \right) \cup \left( \bigcup_{i,j,k,l',j',k'} (\tilde{E}_{ijk} \cap \tilde{E}_{l'j'k'}) \right) \cup \left( \bigcup_{l,i,j,k} (\tilde{D}_l \cap \tilde{E}_{ijk}) \right)$$

$$S^0 = S \cap A$$

(10)

We note that by the foregoing discussion  $S \cap C$  is a system of normally crossing curves in  $S$ , and  $S \cap A$  is a finite set of points in  $S$ . We first consider points on the stratum  $S^2$ .

Case 1.  $p \in S^2$

For a simple point on  $S^2$ , which is the most generic kind of point of  $S$  no blow-ups are permissible. Hence our standard parametrisation of (9) above must be achieved only by a judicious choice of local  $(u, v, w)$  coordinates. This is what we do below

*Proof of case 1*

Let  $p$  be a point on  $S^2$ , and  $U$  be a neighborhood of it not meeting  $C$  (defined above). Choose a holomorphic coordinate  $u$  such that

$$U \cap E = (u = 0)$$

and two other (provisional) coordinates  $v$  and  $w$  such that  $p = (0, 0, 0)$ . Since  $U$  misses  $C$ , and hence all the  $\tilde{D}_i$ , we have

$$z_i = u^{a_i}(\text{unit}) \quad (1 \leq i \leq N),$$

where "unit" means a nowhere vanishing on  $U$  holomorphic function which, by choosing  $U$  to be a small enough polydisc, has a convergent  $(u, v, w)$ -power series expansion on  $U$ . Thus its holomorphic roots or inverse on  $U$  may be taken. By permuting  $z_i$ , and absorbing a holomorphic root of the unit which occurs as the factor in  $u$  into  $u$ , we can assume

$$z_1 = u^{a_1}, z_i = u^{a_i}(\text{unit}) \text{ for } 2 \leq i \leq N, \quad (11)$$

where  $a_N \geq a_i \geq a_{i-1} \dots \geq a_2 \geq a_1$ . Now we decompose each  $z_i$  as

$$z_i = z_{i,1} + z_{i,2} \quad (12)$$

where  $z_{i,1}$  collects the pure  $u$  terms in the  $(u, v, w)$  series expansion of  $z_i$  and  $z_{i,2}$  collects the rest. Clearly both  $z_{i,1}$  and  $z_{i,2}$  are divisible by  $u^{a_i}$ . Since  $a_i \geq a_1$ , we can write

$$z_{i,1} = f_i(u) = f_i(\zeta_1^{1/a_1}) = f'_i(\zeta_1)$$



Since all the powers of  $u$  in  $f_i$  are  $\geq a_1$ , the exponents of  $\zeta_1$  in  $f'_i$  are fractional and  $\geq 1$ . (It is a Puiseux series in  $\zeta_1$ ). Thus (i) of the Proposition 2.2.2 is achieved.

We note that  $z_{i,2} \neq 0$  for  $2 \leq i \leq N$ , because otherwise we would have

$$dz_1 \wedge dz_i \wedge dz_j = dz_1 \wedge dz_{i,2} \wedge dz_j \equiv 0 \text{ on } U$$

for some  $j \neq i$ , contradicting that  $\tilde{E}_{ijk}$  is a surface.

Write  $z_{i,2} = u^{a'_i} h_i$ , where  $a'_i \geq a_1$  and  $h_i$  is indivisible by  $u$ . For simplicity of notation let us drop the prime on  $a'_i$  and call them  $a_i$ , where all  $a_i \geq a_1$  for  $2 \leq i \leq N$ . Again permuting  $z_i$  for  $i \geq 2$ , we may assume that  $a_i \geq a_2$  for all  $2 \leq i \leq N$ . Thus

$$z_i = f_i(u) + u^{a_i} h_i(u, v, w) \quad 2 \leq i \leq N$$

and all  $h_i$  are non-zero, and devoid of constant term since such a term would give a pure  $u$ -term in  $z_{i,2}$  contrary to the definition of  $z_{i,2}$ . Hence  $h_i(0, 0, 0) = 0$  for all  $2 \leq i \leq N$ . In particular,  $h_2$  has no pure  $u$ -terms.

*Claim 1.*  $h_2$  is a local coordinate on  $U$ , after possible shrinking  $U$ .

*Proof of Claim 1.* First we show that  $du \wedge dh_2$  is nowhere vanishing on  $U$ . Since

$$dz_1 \wedge dz_2 = a_1 u^{a_1 + a_2 - 1} du \wedge dh_2$$

and  $dz_1 \wedge dz_2 \wedge dz_i$  for  $i \neq 1, 2$  has no zeros in  $U$  except on the singular set  $E \cap U = (u = 0)$  by the choice of  $U$  (as missing  $\tilde{E}_{12i} \forall i \neq 1, 2$ ), we see that the only possible zeros of  $du \wedge dh_2$  must lie in  $(u = 0)$ . Thus

$$du \wedge dh_2 = u^r \omega,$$

where  $\omega$  is a 2-form nowhere vanishing on  $U$ . To prove the claim, we have to show that  $r = 0$ . Clearly

$$du \wedge dh_2 = \left( \frac{\partial h_2}{\partial v} \right) du \wedge dv + \left( \frac{\partial h_2}{\partial w} \right) du \wedge dw$$

Thus,  $\frac{\partial h_2}{\partial v}$  and  $\frac{\partial h_2}{\partial w}$  are divisible by  $u^r$  by the above two equations. So

$$\frac{\partial h_2}{\partial v} = v^r k_1 \quad \frac{\partial h_2}{\partial w} = u^r k_2,$$

where  $k_i$  are holomorphic functions on  $U$ . Let  $P_v(k_1)$  be the holomorphic  $v$ -primitive of  $k_1$ , got by replacing the monomials  $u^a v^b w^c$  term by term by the monomials  $(1/(b+1))u^a v^{b+1} w^c$  in the  $(u, v, w)$  series expansion of  $k_1$ . Analogously, let  $P_w(k_2)$  be the  $w$ -primitive of  $k_2$ . We therefore get by integrating the last two equations above the following two equations for  $h_2$

$$h_2 = g(u, w) + u^r P_v(k_1) \quad (13)$$

$$h_2 = f(u, v) + u^r P_w(k_2) \quad (14)$$

Looking at the second equation (14) above, we see that since  $h_2$  has no pure  $u$ -terms by (12) every terms in  $f(u, v)$  is of the type  $u^a v^b$  with  $b \geq 1$ . Thus no term of  $f(u, v)$  can occur in  $g(u, w)$  (which, for the same reason has only  $u^a w^c$  terms with  $c \geq 1$ ). So every term of  $f(u, v)$  must figure in the term  $u^r P_v(k_1)$  of the first equation (13). Hence  $f(u, v)$ , and consequently,  $h_2(u, v, w)$  is divisible by  $u^r$ , contradicting that the  $h_i$  were indivisible by  $u$ . So  $r = 0$ .

Now we rename  $h_2$  as the variable  $v_1$ , and forget the original variable  $v$ . We retain  $u$  from before.

Since  $v_1 = h_2$  and  $h_2(\bar{p}) = h_2(u = 0, v = 0, w = 0)$ , the new  $v_1$ -coordinate continues to be 0 at  $p$ . Since the rank of the map

$$(u, v, w) \mapsto (u, v_1)$$

is full all over  $U$ , one can, by shrinking  $U$  and applying the holomorphic implicit function theorem, assume that there exists a third coordinate  $w_1$  such that  $(u, v_1, w_1)$  is a coordinate system on  $U$  and  $w_1(p) = 0$ . This proves the claim.  $\square$

Since  $u$  is unchanged, so is the coordinate function  $z_1 = u^{a_1}$ . We now have

$$z_2 = f_2(u) + u^{a_2} v$$

Expand the coordinate  $z_i$  for  $3 \leq i \leq N$  in a new  $(u, v_1, w_1)$  power series. In the decomposition (12) above, the power series of  $z_{i,1}$  does not change since it contains only pure  $u$  powers. However the  $(u, v_1, w_1)$  power series of  $z_{i,2}$  changes, and may well contain pure  $u$ -terms. But these terms  $u^a$ , since they come from  $z_{i,2}$  are divisible by  $u^{a_i}$ , and  $a_i \geq a_1$  implies they continue to be divisible by  $\zeta_1 = u^{a_1}$ , thus (i) of the proposition continues to be valid. Further, any term in the  $(u, v_1, w_1)$  power series of  $z_i$  which contains  $v_1$  must belong to the  $(u, v_1, w_1)$  power series of  $z_{i,2}$ , and must therefore be divisible by  $u^{a_i}$ , hence  $u^{a_2}$ , hence  $u^{a_2} v_1$ . Let us now drop the subscript on  $v_1$ , and just call it  $v$ .

So we already have that all  $z_i$  are divisible by  $u^{a_1}$ , and all terms containing  $v$  are divisible by  $u^{a_2}$  and hence  $u^{a_2} v$ , and  $a_2 \geq a_1$ . Thus (i) and (ii) of our proposition have been realised.

Now we may make another decomposition

$$z_i = f_i(u, v) + u^{a'_i} h_i(u, v, w_1) \quad (3 \leq i \leq N) \quad (15)$$

where  $f_i(u, v)$  collects all the pure  $(u, v)$  terms in the series expansion of  $z_i$ , and  $h_i(u, v, w_1)$  is indivisible by  $u$ . Note that since  $u^{a'_i} h_i$  is a subseries of (the  $(u, v, w_1)$  series expansion of)  $z_{i,2}$ ,  $a'_i \geq a_i \geq a_2$ . For simplicity of notation we again drop the primes on  $a'_i$  and call them  $a_i$ . Finally note that every term in  $h_i$  for  $3 \leq i \leq N$  must be divisible by  $w_1$  by the definition of the decomposition, so that  $h_i(0, 0, 0) = 0$  for all  $3 \leq i \leq N$ .

By reordering  $3 \leq i \leq N$ , assume  $a_3 \leq a_i$  for all  $3 \leq i \leq N$ . So  $a_3 \geq a_2 \geq a_1 \geq 1$ . Consider

$$dz_1 \wedge dz_2 \wedge dz_3 = (a_1 u^{a_1 + a_2 + a_3 - 1}) \left( \frac{\partial h_3}{\partial w_1} \right) du \wedge dv \wedge dw_1$$

Since  $U \cap \tilde{E}_{123} = \emptyset$ , we have that

$$\frac{\partial h_3}{\partial w_1} = u^r k_1$$

and  $k_1$  is nowhere vanishing holomorphic function on  $U$ . Again this implies

$$h_3 = f(u, v) + u^r P_{w_1}(k_1)$$

But since by definition  $h_3$  contains no pure  $(u-v)$  terms,  $f \equiv 0$  and  $h_3$  is divisible by  $u^r$ , contradicting that  $h_i$  are indivisible by  $u$ . Thus  $r = 0$ , and  $\frac{\partial h_3}{\partial w_1}$  is nowhere vanishing on  $U$ . Thus we see that

$$h_3 = P_{w_1}(k_1) \quad (16)$$

which implies

$$du \wedge dv \wedge dh_3 = \left( \frac{\partial h_3}{\partial w_1} \right) du \wedge dv \wedge dw_1$$

is nowhere vanishing on  $U$ , showing that  $h_3$  may be defined as the new third variable  $w$  and the original  $w_1$  discarded. Since  $h_3(0, 0, 0) = 0$ , the new  $w$  coordinate also vanishes at  $p$ . Thus the new  $(u, v, w)$  coordinate system continues to be centred at  $p$ .

Actually, we have that  $\frac{\partial h_3}{\partial w_1} = k_1$  being a unit on  $U$ , it has a non-zero constant term in its series expansion, so that by (16) above,  $h_3 = w_1(\text{unit})$ , so the new  $w = h_3$  and the old  $w_1$  differ by scaling of a unit.

Finally, note that we have by definition

$$z_3 = f_3(u, v) + u^{a_3} w \quad (17)$$

and since every term containing  $w$  in any  $z_i$  for  $3 \leq i \leq N$  must be contained in the  $(u, v, w)$ -series expansion of  $u^{a_i} h_i$  of the decomposition (15) above (because the pure  $u-v$  part  $f_i(u, v)$  of  $z_i$  remains unchanged under the coordinate change  $(u, v, w_1) \mapsto (u, v, w)$ ), it must be divisible by  $u^{a_i}$ , hence  $u^{a_3}$  and hence  $u^{a_3} w$ . Thus the proposition is proved for the Case 1, once we choose a  $(u, v, w)$  polydisc centred at  $p$  and contained in  $U$ .  $\square$

*Remark 2.2.3.* From the foregoing proof, we clearly have the following interpretation of the integer exponents  $1 \leq a_1 \leq a_2 \leq a_3$ .

$$a_1 = \min_{1 \leq i \leq N} (\text{ord}_E(z_i))$$

$$a_2 = \min_{1 \leq i, j \leq N} (\text{ord}_E(dz_i \wedge dz_j)) - a_1 + 1$$

$$a_3 = \min_{1 \leq i, j, k \leq N} (\text{ord}_E(dz_i \wedge dz_j \wedge dz_k)) - a_1 - a_2 + 1$$

where  $\text{ord}_E$  denotes the order of vanishing along the singular divisor given in  $U$  by

( $u = 0$ ), viz. the largest power of  $u$  factoring out of the respective expressions for  $z_i, dz_i \wedge dz_j, dz_i \wedge dz_j \wedge dz_k$ . This same remark applies to the succeeding cases of simple points, and there are also corresponding interpretations for the double and triple point charts of the various exponents  $a_i, b_i, c_i$  which occur in those situations. In particular, these  $a_i$  remain constant along a fixed component of the singular divisor  $E$ .

*Case 2.*  $p \in S^1$

Next, if  $p$  is a point on the stratum  $S^1$ , it is a smooth point of the curve  $S \cap C$ , and a type-B operation on the smooth (closed) curve component on which  $p$  lies is permissible. Indeed this component of  $S \cap C$  is the smooth intersection of some  $\tilde{D}_i$  or  $\tilde{E}_{ijk}$  with  $E$ . A type-B operation on this curve component  $Q$  on a given simple-point chart centred at  $p$  will result in a simple point and double point charts. There is a *unique relevant* (see Def. 2.2.1 for the definition of ‘relevant’) simple point chart, which will be centred on the new (global, closed) curve of intersection between the new exceptional divisorial component  $\mathbf{CP}(1) \times Q$  (which results from the blow-up of  $Q$ ) and the strict transform of  $\tilde{D}_i$ . *Thus the unique relevant simple-point chart which results from blowing up the curve  $Q$  which locally defines  $S^1$ , which is the only permissible type-B operation on such a simple point chart centred on  $S^1$ , will continue to be centred on (the new)  $S^1$ .* Of course, if we have chosen finitely many charts covering the curve component  $Q$ , all these charts must be subjected to the blow-up, but again the number of resulting relevant simple point charts remains unchanged under this blow-up. There are two subcases.

*Subcase 1 of case 2.* The first and easier one, when  $p \in \tilde{D}_j \cap S$ , and  $p \notin \tilde{D}_i$  for  $i \neq j$ ,  $p \notin \tilde{E}_{ijk} \forall i, j, k$  is given below.

*Proof of subcase 1 of case 2.* Let us choose a neighborhood  $U$  of  $p$  and a system  $(u, v, w)$  of local coordinates in  $U$  centred at  $p$ , just as before, satisfying

- (i)  $S \cap U = (u = 0)$
- (ii)  $\tilde{D}_j \cap U = (v = 0)$

We may later have to modify  $v$  and  $w$  as in Case 1 above. We first note that type-B operations can be done on the (global) closed curve  $\tilde{D}_j \cap S$ , which is given by  $(u = v = 0)$ , viz. the  $w$ -axis in  $U$ . A priori we have, as at the outset of the proof of Case 1 above,

$$\begin{aligned} z_i &= u^{a_i}(\text{unit}) \quad \text{for } i \neq j, 1 \leq i \leq N \\ z_j &= u^{a_j} v^b(\text{unit}) \end{aligned} \tag{18}$$

where “unit” means a holomorphic function nowhere vanishing on  $U$ , whose holomorphic inverses and roots maybe taken, and  $b \geq 1$  is a positive integer.

The initial (and only) obstruction to repeating the proof of Case 1 above is the possibility that for the distinguished index  $j$  above, it may so happen that  $a_j$  may be the least of the  $a_i$ ’s, which means that the factor  $v^b(\text{unit})$ , not being a unit, cannot be scaled away to get  $z_j = u^{a_j}$  (or  $z_1 = u^{a_1}$  after permuting). We remedy this by permissible type-B operations on the  $w$ -axis ( $u = v = 0$ ).

The effect of such a type-B operation on such a simple point chart  $U$  is the following: In the unique new relevant simple point chart  $U'$  created, which continues to be centred around the point  $(0,0,0)$  of  $\pi^*\tilde{D}_j \cap S'$  (where the new set of simple points  $S' =$  the simple points of  $\pi^{-1}(S)$  and  $\pi^*(\tilde{D}_j)$  is the strict transform of  $\tilde{D}_j$ ) the variable  $u, w$  remain unchanged, and  $v$  is replaced by  $uv$ . Since units transform to units under this substitution, the coordinate functions now read on  $U'$  as

$$\begin{aligned} z_i &= u^{a_i}(\text{unit}) \quad \text{for } i \neq j, 1 \leq i \leq N \\ z_j &= u^{a_j+b} v^b(\text{unit}) \end{aligned} \quad (19)$$

Thus the  $u$ -exponent of  $z_j$  gets boosted, and those of the other  $z_i$  remain invariant. So, as in (11) of Case 1 above, we can assume that  $a_j$  is the largest exponent of  $u$  occurring, and thus after rearranging, we may assume

$$\begin{aligned} z_1 &= u^{a_1}(\text{unit}) \\ z_i &= u^{a_i}(\text{unit}) \quad 2 \leq i \leq N-1 \\ z_N &= u^{a_N} v^b(\text{unit}) \end{aligned}$$

where  $a_N \dots \geq a_i \geq a_{i-1} \dots \geq a_1$ .

*Remark 2.2.2.* In the foregoing paragraph,  $a_j = \text{ord}_E(z_j)$ , where  $z_j$  is the unique coordinate whose order of vanishing along  $\tilde{D}_j \geq 1$ . Since orders of vanishing along a divisor remain constant along components of that divisor, we see that boosting  $a_j$  so that  $a_j = \max_i(\text{ord}_E(z_i))$  is simultaneous in *all* the simple point charts centred on the curve  $\tilde{D}_j \cap S$ . In other words, before we localise to a particular  $p$  of  $S^1$  lying on  $\tilde{D}_j \cap S$ , we can *a priori* ensure this maximality of  $a_j$  at all points of the curve  $\tilde{D}_j \cap S$  by blowing up this curve often enough. Then we can localise to an arbitrary point  $p$  of  $S^1$  lying on  $\tilde{D}_j \cap S$ , and the proposition 2.2.2 for the subcase 1 of case 2 discussed above will hold good on a small neighborhood  $U$  of it, and we do not have to worry about the points that are left out when we shrink a given chart.

Now we may proceed as in the proof of Case 1 above. The only additional and redundant fact that is true in this situation is that  $z_N = z_{2,N}$ , and in fact it also follows from the fact that  $U$  avoids all  $\tilde{E}_{ijk}$  that the exponent  $b$  above must be 1, or else  $dz_1 \wedge dz_N$  and hence  $dz_1 \wedge dz_N \wedge dz_j$  for all  $j$  would vanish along  $(v=0)$ .  $\square$

*Proof for subcase 2 of case 2.* Finally we come to last possibility for a point  $p$  on the stratum  $S^1$ . Namely,  $p \in S \cap \tilde{E}_{lmn}$ , such that  $p \notin \tilde{D}_i$  for all  $i$  and  $p \notin \tilde{E}_{ijk}$  for all  $\{i, j, k\} \neq \{l, m, n\}$ . First choose a coordinate neighborhood  $U$  with some coordinates  $(u, v, w)$  centred at  $p$  which satisfies

- (i)  $U \cap \tilde{E}_{ijk} = \emptyset \forall \{i, j, k\} \neq \{l, m, n\}$
- (ii)  $U \cap \tilde{D}_i = \emptyset \forall i$
- (iii)  $U \cap E = (u=0)$
- (iv)  $U \cap \tilde{E} = (w=0)$

$$z_1 = u^{a_1}$$

$$z_i = u^{a_i}(\text{unit}), \quad (20)$$

where  $a_N \geq \dots a_i \geq a_{i-1} \dots \geq a_1$ . The indices  $\{l, m, n\}$  in the statement of the proposition will change, but let us continue to call them  $\{l, m, n\}$  for notational simplicity.

#### DEFINITION 2.2.5

If a 3-form  $\omega$  has the local expression

$$\omega = (u^a w^c \alpha) du \wedge dv \wedge dw$$

where  $\alpha$  is a local unit, we say that  $\text{ord}_u(\omega) = a$  and  $\text{ord}_w(\omega) = c$

*Claim 2.* Without loss of generality we may assume that for the triple  $(l, m, n)$  of the hypothesis, we have

$$\text{ord}_u(dz_1 \wedge dz_m \wedge dz_n) > \text{ord}_u(dz_i \wedge dz_j \wedge dz_k)$$

for all  $(i, j, k)$  such that  $\{i, j, k\} \neq \{l, m, n\}$

*Proof.* We will invoke type-B operations on the  $v$ -axis ( $u = w = 0$ ). Since it is the local germ of the global closed curve  $\tilde{E}_{lmn} \cap E$  these are permissible. Suppose we have the 3-form

$$\omega = (u^a w^c \alpha) du \wedge dv \wedge dw$$

then under a type-B operation on the  $v$ -axis, in the unique new relevant (cf. Def. 2.2.1) simple point chart, where the blow-up map  $\pi$  is given by

$$\pi(u, v, w) = (u, v, uw)$$

we have

$$\pi^* \omega = (u^{a+c+1} w^c \beta) du \wedge dv \wedge dw$$

where  $\beta(u, v, w) = \alpha(u, v, uw)$  is still a local unit.

Thus, for a 3-form  $\omega$  with  $c = \text{ord}_w(\omega) \geq 1$ , we get

$$\text{ord}_u(\pi^* \omega) \geq \text{ord}_u(\omega) + 2$$

whereas for an  $\omega$  with  $c = \text{ord}_w(\omega) = 0$ , we get

$$\text{ord}_u(\pi^* \omega) = \text{ord}_u(\omega) + 1$$

Since by hypothesis and (i), (iv) above,

$$\text{ord}_w(dz_1 \wedge dz_m \wedge dz_n) \geq 1$$

$$\text{ord}_w(dz_i \wedge dz_j \wedge dz_k) = 0 \forall \{i, j, k\} \neq \{l, m, n\} \quad (21)$$

we have the claim in a finite number of type-B operations.  $\square$

*Remark 2.2.6.* We observe here that in line with the Remark 2.2.4 in the proof of subcase 1 of case 2 above, the inequality of the foregoing Claim 2 can be achieved *a priori* all along the curve  $\tilde{E} \cap S$  without localising to any particular point  $p$  on it. The proposition (for subcase 2 of case 2) will hold good in a small neighborhood of an arbitrary point  $p$  of the type in the statement after enough *a priori* type B operations of the curve component  $\tilde{E}_{lmn} \cap S$  on which  $p$  lies.

Decompose, as in the past, for  $2 \leq i \leq N$

$$z_i = z_{i,1} + z_{i,2} = f_i(u) + u^{a_i} w^{c_i} h_i, \quad (22)$$

where  $z_{i,1} = f_i(u)$  collects all the pure  $u$ -terms in the  $(u, v, w)$  power series expansion of  $z_i$ , and  $z_{i,2} = u^{a_i} w^{a_i} h_i$ , (with  $h_i$  indivisible by  $u$  and  $w$ ) is the rest. Thus  $a'_i \geq a_i$  (defined above in (20)), and hence  $\geq a_1$  for all  $2 \leq i \leq N$ . We drop the primes on  $a'_i$  and call them  $a_i$  for notational simplicity. We first claim that

*Claim 3.*

- (a)  $c_i \neq 0$  for at most two indices  $i, 2 \leq i \leq N$
- (b) For any  $i, c_i \neq 0 \Rightarrow c_i = 1$

*Proof.* This is because

$$dz_1 \wedge dz_i = a_1 u^{a_1 + a_2 - 1} w^{c_i - 1} du \wedge (c_i h_i dw + w dh_i)$$

which means that if  $c_i \geq 2$ , we have  $dz_1 \wedge dz_i$  vanishing along  $w = 0$ , which means the 3-forms  $dz_1 \wedge dz_i \wedge dz_k$  will vanish along  $w = 0$  for  $k \neq 1, i$ , which contradicts the hypothesis that there is a unique set  $\{l, m, n\}$  for which  $\tilde{E}_{lmn}$  intersects  $U$ . This proves the claim (b).

To prove (a), we see that

$$dz_1 \wedge dz_i \wedge dz_j = a_1 u^{a_1 + a_i - 1} w^{c_i + c_j - 1} du \wedge (c_j h_j dh_i \wedge dw + c_i h_i dw \wedge dh_j + w dh_i \wedge dh_j)$$

which shows that if both  $c_i, c_j \geq 1$ , the 3-form above vanishes along  $(w = 0)$ . But, by hypothesis this means that  $(1, i, j)$  is a permutation of  $(l, m, n)$ , so this cannot happen for more than a unique pair of indices  $i, j$ , proving (a).  $\square$

By (a) above, there exists a  $j$  (since  $N \geq 4$ !) such that  $c_j = 0$ . Since the permissible type-B operations on the  $v$ -axis have the effect  $u^a \mapsto u^a; u^a w^c \mapsto u^{a+c} w^c$  in the unique relevant simple-point chart which results, we may further assume that if  $c_j = 1$  for some  $j \geq 2$  then  $a_j \neq \min_{i \geq 2} (a_i)$ . Thus by permuting the  $z_i$ 's we can assume that

$$z_2 = f_2(u) + u^{a_2} h_2, \quad (23)$$

where  $h_2$  is indivisible by  $w$  and  $a_2 \leq a_i$  for all  $2 \leq i \leq N$ . Of course by definition  $a_2 \geq a_1$ . Since a constant, or pure  $u$ -term in  $h_2$  would contradict the definition of  $z_{2,2} = u^{a_2} h_2$ , we see that  $h_2(p) = 0$  and  $h_2$  has no pure  $u$  terms.

*Claim 4.* By shrinking  $U$ , we may assume that  $h_2$  is a local coordinate on  $U$ .

*Proof.* As in the proof of the corresponding Claim 1 in Case 1 above, we show that

$du \wedge dh_2$  has no zeroes on  $U$ , and hence, by the implicit function theorem, on a smaller neighborhood centred at  $p$ ,  $(u, v_1, w_1)$  with  $v_1 = h_2$ , is a coordinate system, for some choice of  $w_1$  vanishing at  $p$ .

Any zeros of  $du \wedge dh_2$  outside  $(u=0)$  will be zeros of  $dz_1 \wedge dz_2 \wedge dz_j$  for all  $j \neq 1, 2$ , and there are at least two distinct such  $j$  ( $N \geq 4$ ), we will have a contradiction to the hypothesis that there is a unique 3-form  $dz_1 \wedge dz_m \wedge dz_n$  having zeros outside  $(u=0)$ . Now one repeats the proof of the Claim 1 in Case 1 to show that this is impossible.  $\square$

So, as before, by the implicit function theorem and possibly shrinking  $U$ , we have a coordinate system  $(u, v_1, w_1)$  where  $v_1 = h_2$  (from the Claim 4 above) and  $w_1$  both vanish at  $p$ . To redefine the third coordinate, we proceed as follows:

One decomposes the  $(u, v_1, w_1)$ -power series expansions of  $z_j$  for  $3 \leq j \leq N$

$$z_j = f_j(u, v_1) + u^{a_j} g_j(u, v_1, w_1) \quad (24)$$

where  $f_j$  collects all the terms in the series not involving  $w_1$ , and  $u^{a_j} g_j$  collects the rest, so that by definition all the  $g_j$  are divisible by  $w_1$ . Also we may take out the largest  $u$  power so that  $g_j$  is indivisible by  $u$ . Again

$$dz_1 \wedge dz_2 \wedge dz_j = a_1 u^{a_1 + a_2 + a_j - 1} \left( \frac{\partial g_j}{\partial w_1} \right) du \wedge dv_1 \wedge dw_1 \quad (25)$$

can have zeros lying along  $(w=0) \cup (u=0)$  for any  $3 \leq j \leq N$ . If  $\frac{\partial g_j}{\partial w}$  has any zeros along  $(u=0)$ , we have

$$\frac{\partial g_j}{\partial w} = u^r k_1,$$

where  $k_1$  is indivisible by  $u$ . But this implies, since  $g_j$  has no pure  $u$ - $v$  terms, that

$$g_j = P_w(u^r k_1)$$

is also divisible by  $u^r$ , a contradiction.

Thus  $\frac{\partial g_j}{\partial w_1}$  can have zeros only along  $w=0$ . Furthermore,

$$\text{ord}_u(dz_1 \wedge dz_2 \wedge dz_j) = a_1 + a_2 + a_j - 1$$

If  $\frac{\partial g_j}{\partial w_1}$  does have zeros along  $(w=0)$ , by the hypothesis (i) at the beginning of the proof, we must have  $\{1, 2, j\} = \{l, m, n\}$ , so there is this unique  $j$  for which this happens, and for this  $j$

$$\text{ord}_w(dz_1 \wedge dz_2 \wedge dz_j) \geq 1.$$

However, by the above, combined with Claim 2, we see that

$$a_1 + a_2 + a_j - 1 = \text{ord}_u(dz_1 \wedge dz_2 \wedge dz_j) > \text{ord}_u(dz_1 \wedge dz_2 \wedge dz_k)$$



for all  $k \neq j$ . This implies that

$$a_j > a_k \forall k \neq j.$$

In particular, if  $a_r = \min_{i \neq 1, 2}(a_i)$ , we have that  $r \neq j$  and hence that  $\left(\frac{\partial g_r}{\partial w_1}\right)$  is nowhere vanishing on  $U$ . By relabelling the coordinates we may take  $r = 3$ . Now we may call  $g_3$  as  $w_2$ , and repeat the subsequent steps (see discussion leading upto (16), (17)) of the Propositions 1 and 2 above as before. The proposition follows for the coordinates  $(u, v_1, w_2)$ .

*Case 3.  $p \in S^0$*

Let  $p$  be a simple point in  $S^0$ . There exists at least two other irreducible components from among the  $\{\tilde{D}_i, \tilde{E}_{ijk}\}$  which meet at  $p$ . Since by the lemma in the appendix, all intersections among these components are transverse, and further blow-ups of points and curves along the singular divisor  $E$  on can ensure that these two components meet  $E$  transversely as well, we may find a coordinate system  $(u, v, w)$  centred at  $p = (0, 0, 0)$  so that  $E$  is locally given by  $u = 0$  and these two irreducible components are  $v = 0, w = 0$  respectively. Since now the  $v$  and  $w$  axes are germs of global closed curves on  $E$ , we may perform type B operations along these two axes, and also type A operations on  $p$ , since it is a point on the finite closed set  $S^0$ .

Under the above operations, there will be a *unique relevant* simple point chart centred at the point  $\tilde{p}$  which is the point of intersection of the strict transform of the  $u$ -axis and the (new) exceptional divisor, and will thus again be in the new stratum  $S^0$ . There will of course be other charts, which have been (or will be) separately discussed. Let us now look at the transformation of a monomial under a type A or type B operation in the unique new relevant chart under consideration. They are as follows

- (i) pure  $u$  monomials  $u^a$  remain unchanged under all permissible (i.e. on  $v$  and  $w$  axes) type B and type A operations on  $p = (0, 0, 0)$
- (ii) monomials  $u^a v^b w^c$  replaced by  $u^{a+b} v^b w^c$  under type B operation of the  $w$ -axis
- (iii) monomials  $u^a v^b w^c$  replaced by  $u^{a+c} v^b w^c$  under type B operation on  $v$ -axis
- (iv) monomials  $u^a v^b w^c$  replaced by  $u^{a+b+c} v^b w^c$  under type A operation of  $p = (0, 0, 0)$

Now expand each  $z_j$  as a holomorphic power series in  $(u, v, w)$ , viz.

$$z_j = u^{a_j} f_j(u, v, w),$$

where  $a_j = \text{ord}_E(z_j)$ . Clearly, there exists a pure  $u^a$  term in the expansion of some  $z_j$ , otherwise each  $z_j$  would vanish along the entire  $u$ -axis ( $v = w = 0$ ), contradicting that  $E = \pi^{-1}(0)$  is locally defined by  $(u = 0)$ . Let the least of these be  $u^{a_1}$  in the expansion of  $z_1$ , say. Thus every pure  $u$ -monomial  $u^a$  dominates  $u^{a_1}$ . A monomial which is not a pure  $u$ -monomial is of the type  $u^a v^b w^c$  with  $b + c \geq 1$ . By (i) and (iv) above,  $(a_1 - 1)$  type A operations on the origin leave the pure  $u$  monomials unchanged, and convert  $u^a v^b w^c$  to  $u^{a+(a_1-1)(b+c)} v^b w^c$  and since  $a \geq 1, b + c \geq 1$ , we have

$$a + (a_1 - 1)(b + c) \geq a + a_1 - 1 \geq a_1$$

\* In the sequel, a monomial  $p = u^a v^b w^c$  dominates a monomial  $q = u^d v^e w^f \Leftrightarrow p$  is divisible by  $q \Leftrightarrow a \geq d, b \geq e, c \geq f$ .

and thus, after some type A operations, all monomials in all  $z_j$  power series dominate  $u^{a_1}$ . This means

$$\begin{aligned} z_1 &= \lambda u^{a_1} + (\text{terms dominating } u^{a_1}), \quad \lambda \neq 0 \\ &= u^{a_1}(\text{local unit}), \\ z_j &= u^{a_j} f_j, \end{aligned}$$

where  $a_j \geq a_1 \forall j = 1, 2, \dots, N$ .

Thus (i) of the Proposition 2.2.2 is achieved, and  $a_1 = \min_j(\text{ord}_E(z_j))$ . By absorbing a holomorphic root of the unit into  $u$ , one may rewrite  $z_1 = \zeta_1 = u^{a_1}$ . This does not change the coordinate hyperplanes ( $u=0$ ), ( $v=0$ ), ( $w=0$ ) or coordinate-axes ( $u=w=0$ ) ( $v$ -axis) and ( $u=v=0$ ) ( $w$ -axis). Now, there certainly exists a monomial  $u^a v$  somewhere in some  $z_j$ , for  $j \geq 2$ . Otherwise, the column  $\frac{\partial z_j}{\partial v}$  would vanish along

the nonzero  $u$ -axis ( $u \neq 0, v=w=0$ ), contradicting that the jacobian is of rank three outside ( $u=0$ ). Let  $a_2$  be the minimum of the  $a$ 's such that  $u^a v$  occurs in the expansion of some  $z_j$ . Let us relabel this  $z_j$  as  $z_2$ . Thus  $u^{a_2} v$  is dominated by every monomial of the type  $u^a v$  occurring anywhere in the power series of any  $z_j$ . This condition persists under further type A operations of  $(0, 0, 0)$ , by (iv) above. However, it may not yet be dominated by every monomial containing  $v$ . A monomial containing  $v$ , but not of the type  $u^a v$  is clearly of the form  $v^a v^b w^c$  with  $a \geq a_1$  (by the above), and  $b+c \geq 2$ .

Now,  $(a_2 - 1)$  type A operations on the origin, by (i) and (iv) above leave pure  $u$ -monomials unchanged, convert  $u^{a_2} v$  to  $u^{2a_2-1} v$ , and the monomials  $u^a v^b w^c$  with  $a \geq a_1$ ,  $b+c \geq 2$  to  $u^{a+(a_2-1)(b+c)} v^b w^c$ . Now note that  $b+c \geq 2$  implies  $a+(a_2-1)(b+c) \geq (2a_2-1)$ , so one can assume without loss of generality that every monomial containing  $v$  dominates  $u^{a_2} v$ .

Type-B operations on the  $v$ -axis, by (i) and (iii) above, will leave pure  $u$ -monomials  $u^a$  and monomials  $u^a v^b$  not containing  $w$  unaffected, and convert monomials  $u^a v^b w^c$  with  $c \geq 1$  that involve  $w$  into  $u^{a+c} v^b w^c$ . Thus, by at most  $(a_2 - 1)$  such type B operations on the  $v$ -axis, we can assume that any monomial  $u^a v^b w^c$  with  $c \geq 1$  involving  $w$  satisfies  $a \geq a_2$ .

Now one has

$$z = \zeta_1 = u^{a_1}, z_2 = f_2(\zeta_1) + u^{a_2} v g + h(u, w),$$

where  $f_2(\zeta_1)$  is a Puiseux series (involving  $\geq 1$  fractional powers of  $\zeta_1$ ) which merely collects the pure  $u$  monomials in the (absolutely convergent) power series expansion of  $z_2$ , and hence itself a holomorphic function of  $u$ . Note  $a_2 \geq a_1$ , and  $g$  is a local unit. (All terms containing a non-zero power of  $v$  have been lumped into  $u^{a_2} v g$ , and  $g$ , being a power series with non-zero constant term, is a local unit). By the fact that any monomial with a non-zero power of  $w$  dominates  $u^{a_2} w$ , we may rewrite  $h = u^{a_2} h'$ , and thus

$$z_2 = f_2(\zeta_1) + u^{a_2}(v g + h'(u, w))$$

Now redefine  $v$  as  $g v + h'(u, w)$ . Powers of  $u$  will remain unchanged under this

substitution, and monomials  $u^a v^b$  will become

$$Cu^a v^b + (\text{monomials in only } u, v \text{ dominating } u^a v^b) \\ + (\text{monomials in } u, v, w)$$

where  $C$  is a constant. Thus we get

$$\begin{aligned} z_1 &= u^{a_1} \\ z_2 &= f_2(\zeta_1) + u^{a_2} v \\ z_i &= f_i(u, v, w) \quad \text{for } 3 \leq i \leq N \end{aligned} \quad (27)$$

where  $a_2 \geq a_1$ , and any monomial  $u^a v^b w^c$  in  $f_i$  (for  $3 \leq i \leq N$ ) with  $b \geq 1$  dominates  $u^{a_2} v$ . (Since new monomials in  $w$  have come into being because of the redefinition of  $v$ , no statement is possible now about monomials  $u^a v^b w^c$  containing  $w$ , other than the fact that for such a monomial,  $a \geq a_2 \geq a_1$ ). Also, the coordinate hyperplane ( $v = 0$ ) has changed, so type B operations on the  $w$ -axis ( $(u = v = 0)$ ) are no longer permissible. The origin remains the same, so type A operations are still permitted, which is all we will need hitherto. Thus (i) and (ii) of the statement of Proposition 2.2.2 have been achieved.

Call  $\zeta_2 = u^{a_2} v$ . For  $3 \leq i \leq N$ , separate

$$z_i = g_i(\zeta_1, \zeta_2) + h_i(u, v, w),$$

where  $g_i$  collects all the terms in the expansion of  $z_i$  not involving  $w$ , and  $h_i$  all the terms that do involve  $w$ . Since the power series of  $z_i$  is absolutely convergent,  $g_i$  is a holomorphic power series in  $(u, v)$ , and  $h_i$  is a holomorphic power series in  $(u, v, w)$ . Since formally,

$$u = \zeta_1^{1/a_1} \quad \text{and} \quad v = \zeta_2 \zeta_1^{-a_2/a_1}$$

$g_i$  can be written as a formal series (with fractional exponents) in terms of  $\zeta_1, \zeta_2$  with the clause that substituting  $\zeta_1 = u^{a_1}$ , and  $\zeta_2 = u^{a_2} v$  produces a holomorphic power series in  $(u, v)$ , and each term of which dominates  $u^{a_1}$ , and dominates  $u^{a_2} v$  (if it involves  $v$ ).

If there did not exist a term  $u^a w$  in any of the  $z_j$  for  $3 \leq j \leq N$ , since  $\frac{\partial z_1}{\partial w}$  and  $\frac{\partial z_2}{\partial w}$

are identically zero, we would have an entire column  $\frac{\partial z_j}{\partial w}$  vanishing along the non-zero

$u$ -axis, contradicting that the jacobian of the resolution map  $\pi$  is non-zero outside  $E = (u = 0)$ . Also such a monomial, by (27) clearly satisfies  $a \geq a_2$ , and also must occur in some  $h_j$  by definition. Again let  $a_3$  be the integer such all monomials  $u^a w$  occurring anywhere in any  $h_j$  for  $3 \leq j \leq N$  satisfy  $a \geq a_3$ . Clearly,  $a_3 \geq a_2$ . Let us, by relabelling, assume that  $u^{a_3} w$  occurs in  $h_3$ , the subseries of  $z_3$ . We still do not know that all monomials involving  $w$  dominate  $u^{a_3} w$ . But this is easy to arrange by type A operations, at most  $(a_3 - 1)$  in number, because this converts  $u^{a_3} w$  to  $u^{2a_3-1} w$ , and a monomial  $u^a w$  to  $u^{a+a_3-1} w$ , so that if  $a \geq a_3$ ,  $a + a_3 - 1 \geq 2a_3 - 1$ , and monomials of the type  $u^a w$ , which all dominated  $u^{a_3} w$  transform to monomials  $u^{a'} w$  which

dominate  $u^{a_3}w$  where  $a'_3 = 2a_3 - 1$ . On the other hand, any monomial involving  $v$  and not of the form  $u^a w$ , is of the kind  $u^a v^b w^c$  with  $a \geq a_2 \geq 1$ ,  $c \geq 1$ ,  $b + c \geq 2$ . This is transformed to  $u^{a+(a_3-1)(b+c)} v^b w^c$  under  $(a_3 - 1)$  type A operations, by (iv) above. But then

$$a + (a_3 - 1)(b + c) \geq a + 2a_3 - 2 \geq 2a_3 - 1$$

so that we may assume without loss of generality that all monomials involving  $v$  dominate  $u^{a_3}w$ . Thus

$$z_3 = f_3(\zeta_1, \zeta_2) + u^{a_3}w(\text{local unit}),$$

where any monomial involving  $w$  anywhere dominates  $u^{a_3}$ , and  $a_3 \geq a_2 \geq a_1 \geq 1$ . This local unit can now be absorbed by redefining  $w$ , so that

$$z_3 = f_3(\zeta_1, \zeta_2) + \zeta_3,$$

where  $\zeta_3$  is defined to be  $u^{a_3}$ . Thus (iii) of Proposition 2.2.2 is proved. Hence the proposition.  $\square$

*Remark 2.2.7.* There is an observation to be made about the  $f_i$  above. We claim that the (formal) derivatives of  $f_i$  with respect to the  $\zeta_j$  for  $j = 1, 2, 3$  are also holomorphic functions of  $u, v, w$ , and hence bounded. For, by the expressions for  $u, v, w$  in terms of  $\zeta_j$  in (9)

$$\begin{aligned} \frac{\partial}{\partial \zeta_1} &= \frac{1}{a_1} u^{1-a_1} \frac{\partial}{\partial u} - \frac{a_2}{a_1} u^{-a_1} v \frac{\partial}{\partial v} - \frac{a_3}{a_1} u^{-a_1} w \frac{\partial}{\partial w} \\ \frac{\partial}{\partial \zeta_2} &= u^{-a_2} \frac{\partial}{\partial v} \\ \frac{\partial}{\partial \zeta_3} &= u^{-a_3} \frac{\partial}{\partial w} \end{aligned}$$

Now by (i) of the previous Proposition 2.2.2,  $f_i = u^{a_1} h_i$ , for  $2 \leq i \leq N$  (every monomial everywhere dominates  $u^{a_1}$ ), where  $h_i$  are holomorphic functions. From this and the above expression for  $\frac{\partial}{\partial \zeta_1}$ , it follows that  $\frac{\partial f_i}{\partial \zeta_1}$  is holomorphic. Similarly,  $f_i$  may be written, in view of (ii), (iii) of the previous proposition 2.2.2 as

$$f_i(u, v, w) = h_i(u) + u^{a_2} g_i(u, v, w)$$

(every monomial with a nonzero  $v$  or  $w$  power is divisible by  $u^{a_2}$ ), which implies that the expression above for  $\frac{\partial}{\partial \zeta_2}$  that  $\frac{\partial f_i}{\partial \zeta_2}$  is holomorphic for  $2 \leq i \leq N$ . Note it is zero for  $i = 2$ . Finally writing

$$f_i = p_i(u, v) + u^{a_3} q_i(u, v, w).$$

By (iii) of the previous proposition, every monomial containing a non-zero power of  $w$  is divisible by  $u^{a_3}$ , and applying the expression above for  $\frac{\partial}{\partial \zeta_3}$ , the claim is established.

### PROPOSITION 2.2.8

In standard form, in a neighborhood  $U$  centred at a double point  $p = (0, 0, 0)$  on the singular divisor  $E = \pi^{-1}(0)$  (defined on  $U$  by  $(uv = 0)$ , set-theoretically), the coordinate functions  $z_i$ , after rearrangement, rescaling, and sufficiently many type A, and permissible type B operations, are given by the standard form

$$\begin{aligned} z_1 &= \zeta_1 \\ z_2 &= f_2(\zeta_1) + \zeta_2 \\ z_3 &= f_3(\zeta_1, \zeta_2) + \zeta_3 \\ z_i &= f_i(\zeta_1, \zeta_2, \zeta_3) \quad (4 \leq i \leq N) \end{aligned} \quad (29)$$

where  $\zeta_1 = u^{a_1} v^{b_1}$ ,  $\zeta_2 = v^{a_2} v^{b_2}$ ,  $\zeta_3 = u^{a_3} v^{b_3} w$ ,  $a_3 \geq a_2 \geq a_1 \geq 1$ ,  $b_3 \geq b_2 \geq b_1 \geq 1$ ,  $a_1 b_2 - a_2 b_1 \neq 0$ . Further,

- (i) all monomials  $u^a v^b w^c$  occurring in the expansions of  $f_i$  (in terms of  $u, v, w$ ), dominate  $\zeta_1$ .
- (ii) all monomials  $u^a v^b$  occurring in the expansions of any  $f_i$  which obey  $ab_1 - a_1 b \neq 0$  dominate  $\zeta_2$ .
- (iii) all monomials containing a nonzero power of  $w$  dominate  $\zeta_3$ .
- (iv) Finally, the  $\zeta_i$ -derivatives of  $f_j$ , from the conditions imposed on  $a, b, c$  above, are bounded holomorphic functions in the small neighborhood  $U$  under consideration.

*Proof.* The double points lie on the system of curves where two irreducible components of  $E$  the exceptional divisor intersect. We shall always assume these to be given by  $(u = 0)$  and  $(v = 0)$  in a local coordinate system  $(u, v, w)$ . If we let  $B$  denote the curve of double points, then as in (10), we have two strata in  $B$ , viz.,

$$\begin{aligned} B^1 &= B \cap (C - A) \\ B^0 &= B \cap A \end{aligned} \quad (30)$$

where  $A$  and  $C$  have the same meanings as in (10).

*Case 1.*  $p \in B^0$

This is the case when a third component of  $C$ , say  $(w = 0)$  passes through the point  $p = (0, 0, 0)$  under consideration. Since in this case the local  $u, v, w$ -axes are local germs of globally defined closed curves of intersection of the components of  $C$  and/or  $E$  (given locally by  $(u = 0)$ ,  $(v = 0)$ ,  $(w = 0)$ ) type A operations on  $p = (0, 0, 0)$  and type B operations on all the axes are permissible. Under further type A or B operations a double point chart centred at a point in  $B^0$  creates new relevant (see Def. 2.2.1 for the definition of relevant) double point chart (or charts), which are centred at a point of the (new) stratum  $B^0$ , together with other kinds of charts which have been (or will be) discussed separately. More precisely,

- (i) Under a type B operation on the  $u$ -axis ( $v = w = 0$ ), we get a unique new relevant double-point chart centred at the origin, which is a point of the new stratum  $B^0$ .  $u$  and  $v$  remain unchanged,  $w$  is replaced by  $vw$ .

- (ii) Under a type B operation on the  $v$ -axis ( $u = w = 0$ ), we get a unique relevant new double-point chart centred at the origin, which is a point of the new stratum  $B^0$ .  $u$  and  $v$  remain unchanged,  $w$  is replaced by  $uw$ .
- (iii) Under a type B operation on the  $w$ -axis, ( $u = v = 0$ ), we get two new relevant double point charts centred at the origin, which are both points of the new  $B^0$ .  $w$  remains unchanged in both these charts.  $u$  remains unchanged and  $v$  is replaced by  $uv$  in one, whereas  $v$  remains unchanged and  $u$  is replaced by  $uv$  in the other.
- (iv) Under a type A operation at the origin, again there are two new relevant double point charts centred at the origin, which are both points of the new  $B^0$ . In one,  $u$  remains unchanged,  $v$  is replaced by  $uv$ ,  $w$  is replaced by  $uw$ . In the other,  $v$  remains unchanged,  $u$  is replaced by  $uv$  and  $w$  is replaced by  $vw$ .

In terms of lists, we see that under all the above, the  $c_j$  column remains unchanged. Thus, we make the following.

**Remark 2.2.9** If we start with an arbitrary list  $(a_j, b_j, c_j)_{j=1}^N$  which has  $c_j = 0$  for some  $j$ , and appeal to Lemma 2.1.5 to arrange them, so that after relabelling,  $i \leq j \Rightarrow a_i \leq a_j, b_i \leq b_j, c_i \leq c_j$ , then in the charts discussed in (i) through (iv) which we are interested in,  $c_1 = 0$ . This is because the  $c_j$  column remained unaffected throughout, and hence always has a 0 occurring somewhere, and after the ordering achieved by that lemma, since  $c_1 = \min_i(c_i)$ ,  $c_1$  must be zero.

By this remark, we may assume that, after relabelling the  $z_j$ , in the new relevant charts we are interested in,

$$z_i = u^{a_i} v^{b_i} w^{c_i} f_i(u, v, w) \quad (31)$$

where  $c_1 = 0$ , and  $a_i \leq a_j, b_i \leq b_j, c_i \leq c_j$  for  $i \leq j$ . By the discussion (loc. cit.) all the  $f_i$  are local units.

Hence  $z_1 = u^{a_1} v^{b_1} f_1$ , where  $f_1$  is a local unit. By redefining  $u$  after multiplying it with a unit (the  $a_1$ -th root of  $f$ ), we can take

$$z_1 = u^{a_1} v^{b_1} = \zeta_1$$

Since the redefinition of  $u$  only altered the  $f_i$ 's for  $2 \leq i \leq N$  by units, and  $a_i \geq a_1$  and  $b_i \geq b_1$  for all  $i$ , (i) of the Proposition 2.2.8 is achieved.

Now decompose each  $z_j$  as

$$z_j = f_j(\zeta_1) + z_{j,1},$$

where  $f_j$  collects all the monomials  $u^a v^b$  in the expansion of  $z_j$  with  $ab_1 - a_1 b = 0$ , and  $z_{j,1}$  is the rest. The jacobian of the resolution map being non-zero outside  $E = (uv = 0)$  implies that  $z_{j,1} \neq 0$  for some  $j \geq 2$ . The same reasoning shows that all the  $z_{j,1}$ 's cannot vanish along  $(w = 0)$ , for otherwise along  $(w = 0, uv \neq 0)$ , all the  $dz_j$  would be multiples of  $dz_1$ , contradicting that the resolution map is biholomorphic on  $(w = 0) - E$ . Thus

$$z_{j,1} = u^{a_j} v^{b_j} h_j \quad (32)$$

with  $a_j \geq a_1, b_j \geq b_1$ , (there  $a_j$ 's and  $b_j$ 's are not the ones in (31) above), and some  $h_j$  is not divisible by  $w$ . We can factor out the largest power  $w^{c_j}$  from each  $h_j$ , apply

Lemma 2.1.5 to the list  $(a_j, b_j, c_j)$ , noting that some of the  $c_j$ 's are zero. Relabelling, we may assume that, in the resulting relevant double point charts as discussed in (i) through (iv) above, and by the remark 2.2.9,

$$z_{2,1} = u^{a_2} v^{b_2} h_2(u, v, w), \quad (33)$$

where  $a_j \geq a_2$ ,  $b_j \geq b_2$  (which are defined above in (32)), for  $2 \leq j \leq N$ , and  $h_2$  is not divisible by  $w$ .

Now  $h_2$  may not be a local unit. Let us write

$$h_2(u, v, w) = d_0(u, v) + d_1(u, v)w + \cdots d_i(u, v)w^i + \cdots, \quad (34)$$

where we know that  $d_0(u, v)$  is not identically 0. By enough type B (allowed on all three axes) operations, we would like to arrange that in the resultant double point charts of (i) through (iv) above,  $h_2(u, v, w)$  is of the form  $u^a v^b$  (local unit). We do this in the next sublemma.

*Sublemma 2.2.10. Let  $h(u, v, w)$  be given by a power series (34) above. After enough type B operations, one may assume that one has*

$$h(u, v, w) = u^a v^b (\text{local unit})$$

*in all the relevant (see 2.2.1 for the definition of relevant) double point charts discussed in (i) through (iii) above.*

*Proof.* As in Lemma 2.1 of [5], after enough type B operations on  $(u = v = 0)$ , the  $w$ -axis, we can ensure that  $d_0(u, v) = u^a v^b$  (local unit) in all resulting relevant double point charts. Now we need to blow up the  $u, v$ -axes in the new double point charts. We need to do type B blow-ups of the  $u$  and  $v$  axes, so that in the resultant double point charts,  $d_i(u, v)$  is divisible by  $u^a v^b$  for all  $i$ . This is arranged as follows. Blowing up the  $v$ -axis by a type B operation means that in the new relevant double point chart (the other chart will be a triple point chart, see (ii) above), we are making the substitutions

$$w \rightarrow uw, \quad v \rightarrow v, \quad u \rightarrow u,$$

so that

$$\begin{aligned} d_i(u, v)w^i &\rightarrow u^i d_i(u, v)w^i \quad i \geq 1 \\ d_0(u, v) &\rightarrow d_0(u, v). \end{aligned}$$

Similarly, a type B operation on the  $u$ -axis would mean (see (i) above)

$$\begin{aligned} d_i(u, v)w^i &\rightarrow v^i d_i(u, v)w^i \quad i \geq 1 \\ d_0(u, v) &\rightarrow d_0(u, v). \end{aligned}$$

Thus repeating  $v$ -axis blow-ups at most  $a$  times, and  $u$ -axis blow-ups at most  $b$  times, we will have

$$z_{2,1} = u^{a_2} v^{b_2} (\text{local unit})$$

and type B operations preserve the fact that  $z_{j,1}$  dominates  $u^{a_2} v^{b_2}$  for all  $j \geq 2$ . If we call this local unit  $h$ , by making a change  $u \rightarrow u\delta$ ,  $v \rightarrow v\gamma$ , where  $\delta$  and  $\gamma$  are local units satisfying  $\delta^{a_1} \gamma^{b_1} = 1$ , and  $\delta^{a_2} \gamma^{b_2} = h^{-1}$ , (which is possible since  $a_1 b_2 - a_2 b_1 \neq 0$ ) we have

$$z_2 = f_2(\zeta_1) + \zeta_2$$

where  $\zeta_i = u^{a_i} v^{b_i}$ , for  $i = 1, 2$ , and thus we achieve (ii) of the Proposition 2.2.8.

It remains to achieve (iii) of the proposition. Now decompose

$$z_j = f_j(\zeta_1, \zeta_2) + z_{j,2}$$

for  $j \geq 3$ , where  $f_j$  collects all monomials devoid of  $w$ , and hence may be expressed as a formal series with rational exponents of  $\zeta_1, \zeta_2$ , (since  $a_1 b_2 - a_2 b_1 \neq 0$ ,  $u$  and  $v$  can be expressed as monomials in rational powers of  $\zeta_1, \zeta_2$ ), and  $z_{j,2}$  collects all the terms containing  $w$ . Note that  $z_{j,2}$  are all divisible by  $\zeta_2 = u^{a_2} v^{b_2}$ , because  $z_{j,2}$  is a subseries of the  $z_{j,1}$  above which were all divisible by  $u^{a_2} v^{b_2}$ . (E.g., for those  $z_j$  which vanish along ( $w = 0$ ), clearly  $z_j = z_{j,2}$ ).

A first power of  $w$  occurs somewhere in some  $z_{j,2}$ , otherwise the column  $\frac{\partial z_j}{\partial w} = \frac{\partial z_{j,2}}{\partial w}$  of the jacobian of the resolution map would vanish along ( $w = 0$ ), which is not part of  $E$ , contradicting that the map is biholomorphic outside  $E$ . By appeal to Lemma 2.1.5, and the Sublemma 2.2.10 above, and relabelling, we may write

$$z_{3,2} = u^{a_3} v^{b_3} w (\text{local unit})$$

and so that  $z_{j,2}$  dominates  $u^{a_3} v^{b_3} w$  for all  $j \geq 3$ . We redefine  $w$  by absorbing the local unit into it, and write  $\zeta_3 = u^{a_3} v^{b_3} w$ . This proves (iii) of the Proposition 2.2.8, and we're done in the Case 1.  $\square$

*Case 2.  $p \in B^1$*

The proof of this case follows by adapting the above case in a manner analogous to the for Subcase 1 of Case 2 in the proof of 2.2.2. The third variable  $w$ , which is not canonically defined, has to be carefully chosen, as we did there. We omit the straightforward details.

The last statement (iv) of the proposition, about the  $\zeta_i$  derivatives of the  $f_j$  follows exactly as in the Remark 2.2.7.

### 2.23 The analysis of triple-point charts

The main proposition is the obvious analogue of Propositions 2.2.2 and 2.2.8. Namely

#### PROPOSITION 2.2.11

*In a small neighbourhood  $U$  of a triple point  $p$ , after enough type A and B operations, there exists a  $(u, v, w)$  coordinate-system centred at  $p$ , so that the exceptional divisor*



$E$  is defined in  $U$  by  $(uvw = 0)$  set-theoretically, and after relabelling and rescaling, the coordinate functions  $z_j$  are given by

$$\begin{aligned} z_1 &= \zeta_1 \\ z_2 &= f_2(\zeta_1) + \zeta_2 \\ z_3 &= f_3(\zeta_1, \zeta_2) + \zeta_3 \\ z_i &= f_i(\zeta_1, \zeta_2, \zeta_3) \quad (4 \leq i \leq N) \end{aligned} \quad (35)$$

where  $\zeta_i = u^{a_i} v^{b_i} w^{c_i}$  for  $i = 1, 2, 3$ , with the determinant

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$

not equal to 0. Further  $f_i$ , when expressed in terms of  $u, v, w$  are holomorphic functions of  $u, v, w$ , and consist of monomials  $u^a v^b w^c$ , which

- (i) always dominate  $\zeta_1$
- (ii) dominate  $\zeta_2$  if  $(a, b, c)$  is linearly independent of  $(a_1, b_1, c_1)$ .
- (iii) dominate  $\zeta_3$  if  $(a, b, c)$  is linearly independent of both  $(a_1, b_1, c_1)$  and  $(a_2, b_2, c_2)$ .
- (iv) Finally, the  $\zeta_i$ -derivatives of  $f_j$ , from the conditions imposed on  $a, b, c$  above, are bounded holomorphic functions in the small neighborhood  $U$  under consideration.

*Proof.* The triple points are finitely many in number. By type A and B operations, we can ensure that none of the components of  $C$  (see (10) for the definition of  $C$ ) go through our triple point  $p$ . We choose a coordinate system  $(u, v, w)$  which is centred at  $p$ , so that the three irreducible components of  $E$  are defined set-theoretically in  $U$  by  $(uvw = 0)$ . Thus at the outset

$$z_i = u^{a_i} v^{b_i} w^{c_i} (\text{local unit}) \quad (36)$$

By using Lemma 2.1.5, one can assume  $a_i \geq a_1, b_i \geq b_1, c_i \geq c_1$ . Now absorb the root of a unit into  $u$ , so that  $z_1 = \zeta_1 = u^{a_1} v^{b_1} w^{c_1}$  and all monomials everywhere dominate  $\zeta_1$ , thus achieving (i) of our proposition. Now separate

$$z_i = f_i(\zeta_1) + z_{i,1} \quad (2 \leq i \leq N),$$

where  $f_i$  collects all monomials  $u^a v^b w^c$  in the  $(u, v, w)$  power series expansion of  $z_i$  such that  $(a, b, c)$  is linearly dependent on (i.e. a rational multiple of)  $(a_1, b_1, c_1)$  (so may be written as a Puiseux series  $f_i(\zeta_1)$  with  $\geq 1$  fractional exponents in  $\zeta_1$ , and  $z_{i,1}$  collects the rest of the series of  $z_i$ . Again, type A and B operations ensure that

$$z_{i,1} = u^{a_i} v^{b_i} w^{c_i} (\text{local unit}) \quad (2 \leq i \leq N)$$

(where these  $a_i$ 's,  $b_i$ 's and  $c_i$ 's are different from the ones in (36) above) and also that  $a_i \geq a_2, b_i \geq b_2, c_i \geq c_2$  for all  $2 \leq i \leq N$ . Again we absorb a root of the local-unit factor of  $z_{2,1}$  into  $v$ , call  $\zeta_2 = u^{a_2} v^{b_2} w^{c_2}$ , and conclude that

$$z_2 = f_2(\zeta_1) + \zeta_2$$

and the condition (ii) in the statement of the proposition is met. Now we repeat the entire argument after separating

$$z_i = f_i(\zeta_1, \zeta_2) + z_{i,2} \quad (3 \leq i \leq N),$$

where  $f_i$  collects all monomials  $u^a v^b w^c$  in the  $(u, v, w)$  power series expansion of  $z_i$  such that  $(a, b, c)$  is linearly dependent on the pair of triples  $(a_1, b_1, c_1), (a_2, b_2, c_2)$  (so may be written as a formal series  $f_i(\zeta_1, \zeta_2)$  in fractional exponents of  $\zeta_1, \zeta_2$ ) and  $z_{i,2}$  collects the rest of the series of  $z_i$ . Now repeat the reasoning used for  $z_{i,1}$  above to deal with  $z_{i,2}$ , and the part (iii) of the proposition follows. The last claim in the statement follows by a repeat of the proof of Remark 2.2.7.

#### 2.2.4 Final remarks

There is a “stability” about the parametrisations of (9), (29) and (35) above.

For example, we take a simple point polydisc chart  $U$  centred at  $(u, v, w) = (0, 0, 0)$  with the parametrization of Proposition 2.2.2 above, and recenter the coordinates at another point  $(0, s, t)$  in  $U \cap E$ . This means introducing new coordinates

$$\begin{aligned} u &= u, \\ v_1 &= v - s, \\ w_1 &= w - t. \end{aligned} \tag{37}$$

It is easy to see that the parametrisation of the Proposition 2.2.2 persists in the new coordinates  $(u, v_1, w_1)$ , with no change in indices, exponents, and only a change in the functions  $f_2(u)$ ,  $f_3(u, v)$  and  $f_i(u, v, w)$ . But all the assertions of that proposition remain valid. For,

$$\begin{aligned} z_1 &= u^{a_1}, \\ z_2 &= f_2(u) + u^{a_2} v = (f_2(u) + s u^{a_2}) + u^{a_2} v_1, \\ &= f'_2(u) + u^{a_2} v_1 \\ z_3 &= f_3(u, v) + u^{a_3} w, \\ &= (f_3(u, v_1 + s) + t u^{a_3}) + u^{a_3} w_1, \\ &= f'_3(u, v_1) + u^{a_3} w_1. \end{aligned}$$

The rest of the equations too will transform, viz.,

$$z_i = f_i(u, v, w) = f'_i(u, v_1, w_1) \quad (3 \leq i \leq N).$$

Since  $u$  has not changed, every term in every  $z_i$  is still divisible by  $u^{a_1}$ . Any term  $u^a v_1^b w_1^c$  with  $b \geq 1$  involving  $v_1$  in  $f'_i$  for  $(2 \leq i \leq N)$  transforms to the term

$$u^a v^b w^c + (\text{lower order in } v)$$

in the original  $f_i$ , so that by induction on  $b$  and the fact that a term in  $f_i$  involving  $v$  is divisible by  $u^{a_2}$ , one sees that  $a \geq a_2$ , and the term is therefore divisible by  $u^{a_2} v_1$ . Similarly, any term in  $f'_i$  for  $(3 \leq i \leq N)$  involving  $w_1$  is also divisible by  $u^{a_3} w_1$ .

Similar considerations apply to double and triple point charts. For instance, let us take the parametrisation of a neighborhood of a double point  $p$  as asserted by the equations (29) in 2.2.8. If we reset the origin (say) at the point  $(0, x, 0)$ , and consider a small neighborhood of it, we will first be changing coordinates ( $u \rightarrow u$ ,  $v \rightarrow v + x$ ,  $w \rightarrow w$ ) to get

$$z_1 = u^{a_1}(x + v)^{b_1} = u_1^{a_1}$$

where

$$u_1 = u(x + v)^{b_1/a_1}$$

$$\begin{aligned} z_2 &= f_2(\zeta_1) + u^{a_2}(x + v)^{b_2} \\ &= f'_2(u_1) + u_1^{a_2}(x + v)^{b_2 - (a_2 b_1/a_1)} \\ &= f'_2(u_1) + C u_1^{a_2} + u_1^{a_2} v_1 \end{aligned}$$

where

$$\begin{aligned} C &= x^{b_2 - (a_2 b_1/a_1)}, \quad \text{and } v_1 = (x + v)^{b_2 - (a_2 b_1/a_1)} - C \\ &= f''_2(u_1) + u_1^{a_2} v_1 \end{aligned} \tag{38}$$

where  $f_2(\zeta_1) = f_2(z_1) = f_2(u_1^{a_1}) = f'_2(u_1)$  in the third line above. Finally, since  $\zeta_1 = u_1^{a_1}$  and  $\zeta_2 = u_1^{a_2} v_1$  from the above,  $f_3(\zeta_1, \zeta_2)$  is  $f'_3(u_1, v_1)$ , and the expression

$$\begin{aligned} \zeta_3 &= u_1^{a_3}(x + v)^{b_3 - (a_3 b_1/a_1)} w \\ &= u_1^{a_3} w_1 \\ \Rightarrow z_3 &= f'_3(u_1, v_1) + u_1^{a_3} w_1 \end{aligned} \tag{39}$$

where  $w_1 = w(x + v)^{b_3 - (a_3 b_1/a_1)}$ . This is exactly the parametrisation (9) of Proposition 2.2.2 above.

### 2.3 Local models of the metric

In this section we analyse the induced Fubini metric in small neighborhoods of simple, double and triple points. The main fact is

#### PROPOSITION 2.3.1

*In a small neighborhood of a simple, double or triple point the induced Fubini-Study metric is quasi-isometric to  $\sum_{i=1}^3 d\zeta_i d\bar{\zeta}_i$ , after the coordinate functions  $z_i$  have been brought to the standard forms of (9), (29), (35) respectively, (q.v. for definition of  $\zeta_i$ ).*

*Proof.* Let

$$\alpha \frac{\partial}{\partial u} + \beta \frac{\partial}{\partial v} + \gamma \frac{\partial}{\partial w}$$

be a tangent vector. Then the length of this vector in the induced Fubini Study metric (quasi-isometric to the Euclidean metric in a small affine chart in  $(z_i)$ -space centred at the origin which is the isolated singularity of our three fold germ) is given by

$$\sum_{i=1}^N \left| \left( \frac{\partial z_i}{\partial u} \right) \alpha + \left( \frac{\partial z_i}{\partial v} \right) \beta + \left( \frac{\partial z_i}{\partial w} \right) \gamma \right|^2$$

For convenience let us denote the expression

$$\left(\frac{\partial z_i}{\partial u}\right)\alpha + \left(\frac{\partial z_i}{\partial v}\right)\beta + \left(\frac{\partial z_i}{\partial w}\right)\gamma = A_i$$

and the expression

$$\left(\frac{\partial \zeta_i}{\partial u}\right)\alpha + \left(\frac{\partial \zeta_i}{\partial v}\right)\beta + \left(\frac{\partial \zeta_i}{\partial w}\right)\gamma = B_i$$

We have to show that the quantity

$$\frac{\sum_{i=1}^N |A_i|^2}{\sum_{i=1}^3 |B_i|^2} \quad (40)$$

is bounded above and below by (strictly) positive constants. To show it is bounded above, it suffices to show that

$$\frac{|A_i|^2}{\sum_{i=1}^3 |B_i|^2} \quad (41)$$

is bounded above for all  $i = 1, 2, \dots, N$ . Now from (9), (29), (35), this is clear for  $i = 1$  since  $A_1 = B_1$ . For  $i = 2$ ,

$$\begin{aligned} A_2 &= \left(\frac{\partial f_2}{\partial \zeta_1}\right) B_1 + B_2 \\ \Rightarrow |A_2|^2 &\leq 2 \left[ \left| \frac{\partial f_2}{\partial \zeta_1} \right|^2 |B_1|^2 + |B_2|^2 \right] \end{aligned} \quad (42)$$

by the Cauchy-Schwartz inequality. But, by the Remark 2.2.7 and the statement (iv) of the Propositions 2.2.8, 2.2.11 of the last section,  $\left| \frac{\partial f_2}{\partial \zeta_1} \right|$  is bounded above in a small neighborhood, so that (41) is bounded above for  $i = 2$ . For  $i = 3$ , we have

$$A_3 = \sum_{j=1}^2 \left(\frac{\partial f_3}{\partial \zeta_j}\right) B_j + B_3 \quad (43)$$

and for  $4 \leq i \leq N$

$$A_i = \sum_{j=1}^3 \left(\frac{\partial f_i}{\partial \zeta_j}\right) B_j \quad (44)$$

So that we have, again by the Cauchy-Schwartz inequality, for  $i = 3$

$$|A_3|^2 \leq 3 \left( \sum_{j=1}^2 \left| \frac{\partial f_3}{\partial \zeta_j} \right|^2 |B_j|^2 + |B_3|^2 \right)$$

and for  $4 \leq i \leq N$

$$|A_i|^2 \leq 3 \sum_{j=1}^3 \left| \frac{\partial f_i}{\partial \zeta_j} \right|^2 |B_j|^2$$

But, by Remark 2.2.7 and part (iv) of the Propositions 2.2.8 and 2.2.11 of the last section,  $\left| \frac{\partial f_i}{\partial \zeta_j} \right|$  is bounded above in a small neighborhood, so that (41) is bounded above for  $i \geq 3$ .

To show that (40) is bounded below, it suffices to show that

$$\frac{\sum_{i=1}^3 |A_i|^2}{\sum_{i=1}^3 |B_i|^2} \quad (45)$$

is bounded below. Now from the first equation of (42), and (43) above and the Cauchy-Schwartz inequality we have, for  $\varepsilon, \delta < 1$

$$|A_2|^2 \geq \varepsilon \left[ - \left| \frac{\partial f_2}{\partial \zeta_1} \right|^2 |B_1|^2 + \frac{1}{2} |B_2|^2 \right], \quad (46)$$

$$|A_3|^2 \geq \delta \left[ - \sum_{j=1}^2 \left| \frac{\partial f_3}{\partial \zeta_j} \right|^2 |B_j|^2 + \frac{1}{3} |B_3|^2 \right]. \quad (47)$$

We may choose the constants  $\varepsilon$  and  $\delta$  so that in a small  $(u, v, w)$  neighborhood, the following inequalities are satisfied

$$\varepsilon \left| \frac{\partial f_2}{\partial \zeta_1} \right|^2 \leq \frac{1}{4}$$

$$\delta \left| \frac{\partial f_3}{\partial \zeta_1} \right|^2 \leq \frac{1}{4}$$

$$\delta \left| \frac{\partial f_3}{\partial \zeta_2} \right|^2 \leq \frac{\varepsilon}{4}$$

Note that these three inequalities are possible by boundedness of the partial derivative of  $f_i$  for  $i = 2, 3$  with respect to  $\zeta_j$  for  $j = 1, 2, 3$ .

By adding (46), (47) and  $|A_1|^2 = |B_1|^2$ , we find that

$$\begin{aligned} \sum_{i=1}^3 |A_i|^2 &\geq \frac{1}{2} |B_1|^2 + \frac{\varepsilon}{4} |B_2|^2 + \frac{\delta}{3} |B_3|^2 \\ &\geq K \left( \sum_{i=1}^3 |B_i|^2 \right) \end{aligned}$$

where  $K = \min \left( \frac{1}{2}, \frac{\varepsilon}{4}, \frac{\delta}{3} \right)$ . This proves (45), and hence (40) is bounded below by  $K$ .

**Lemma 2.3.2** *In a small  $(u, v, w)$ -polydisk neighborhood  $W$  centred at a point on the exceptional divisor  $E$ , the coordinates  $r_1 = |\zeta_1|$  and  $r = \left( \sum_{i=1}^N |z_i|^2 \right)^{1/2}$  can be interchanged without altering the quasiisometry class of the metric models obtained in the Proposition 2.3.1.*

*Proof.* The coordinate functions  $z_i$  are given by the equations (9), (29), (35). (Notation: Here  $z_1 = \zeta_1$  and let us denote  $r_1 = |z_1| = |\zeta_1|$ ,  $r_i = |z_i|$ ,  $\theta_i = \arg z_i$ .) From those equations it is clear that

$$\frac{z_i}{z_1} = \frac{z_i}{\zeta_1}$$

is a holomorphic function on  $W$ , and hence bounded above there, for  $i = 1, N$ . This implies that on the small  $(u, v, w)$ -polydisk neighborhood  $W$  as stated,

$$1 \leq \frac{r^2}{r_1^2} = \sum_{i=1}^N \left( \frac{r_i}{r_1} \right)^2 \leq C \quad (48)$$

for some positive constant  $C$ .

The original Fubini-Study metric is

$$\sum_{i=1}^N dz_i d\bar{z}_i = \sum_{i=1}^N (dr_i^2 + r_i^2 d\theta_i^2) \quad (49)$$

where  $r_i$  and  $\theta_i$  are as defined above. Now consider the metric got by replacing  $r_1$  by  $r = (\sum_i |z_i|^2)^{1/2}$ . For a constant  $K$  (to be specified later), this new metric is clearly quasi-isometric to

$$dr^2 + r^2 d\theta_1^2 + \sum_{i=2}^N (K dr_i^2 + r_i^2 d\theta_i^2). \quad (50)$$

We claim that on the small polydisk  $W$  this is quasi-isometric to (49) above. Since

$$r^2 = \sum_i r_i^2 \quad (51)$$

$$dr = \left( \frac{r_1}{r} \right) dr_1 + \sum_{i=2}^N \left( \frac{r_i}{r} \right) dr_i \quad (52)$$

By the Cauchy-Schwartz inequalities

$$|a + b|^2 \geq \frac{1}{2}|a|^2 - |b|^2$$

and

$$\left| \sum_{j=1}^p a_j \right|^2 \leq p \left( \sum_{j=1}^p |a_j|^2 \right)$$

it follows that we have the following inequalities:

$$\begin{aligned} & \frac{1}{2} \left( \frac{r_1}{r} \right)^2 dr_1^2 - (N-1) \sum_{i=2}^N \left( \frac{r_i}{r} \right)^2 dr_i^2 \\ & \leq dr^2 \leq N \left( \left( \frac{r_1}{r} \right)^2 dr_1^2 + \sum_{i=2}^N \left( \frac{r_i}{r} \right)^2 dr_i^2 \right) \end{aligned} \quad (53)$$

Thus the metric in (50) is bounded above by

$$N \left( \left( \frac{r_1}{r} \right)^2 dr_1^2 \right) + \left( \frac{r}{r_1} \right)^2 r_1^2 d\theta_1^2 + \sum_{i=2}^N \left( K + N \left( \frac{r_i}{r} \right)^2 \right) dr_i^2 + \sum_{i=2}^N r_i^2 d\theta_i^2 \quad (54)$$

and bounded below by

$$\frac{1}{2} \left( \frac{r_1}{r} \right)^2 dr_1^2 + \sum_{i=2}^N \left( K - (N-1) \left( \frac{r_i}{r} \right)^2 \right) dr_i^2 + \sum_{i=1}^N r_i^2 d\theta_i^2 \quad (55)$$

From the inequality (48) above, and the fact that on  $W$  we have  $\frac{r_i}{r} \leq 1$  for all  $i = 1, \dots, N$ , the expressions (54), (55) above are bounded above and below respectively by

$$A \sum_{i=1}^N (dr_i^2 + r_i^2 d\theta_i^2)$$

and

$$B \sum_{i=1}^N (dr_i^2 + r_i^2 d\theta_i^2)$$

for some positive constants  $A$  and  $B$ , provided one chooses  $K$  to satisfy, for example,

$$K \geq (N-1) \left( \frac{r_i}{r} \right)^2 + 1$$

for all  $i = 2, \dots, N$ . But since on  $W_\alpha$  we have  $\frac{r_i}{r} \leq 1$  for all  $i$ , we can simply take  $K = N$  to ensure this all over  $W$

This establishes that (50) is quasi-isometric to (49) □

### 3. Self-adjointness of the Laplacian

#### 3.1 The basic estimate

**Notation 3.1.1** In what follows,  $r_i = |\zeta_i|$ ,  $\theta_i = \arg \zeta_i$  (where the  $\zeta_i$  are defined in (2.2.2, 2.2.8, 2.2.11)). Also  $u = \rho e^{i\phi}$ ,  $v = \tau e^{i\psi}$ ,  $w = \sigma e^{i\chi}$ . The sets

$$\{(u, v, w) : 0 \leq \rho < 1, \quad 0 \leq \tau < 1, \quad 0 < r_1 < b < 1\}$$

in the three situations corresponding to simple, double or triple points of the Propositions of the last section will be referred to as  $W_I^b$ ,  $W_{II}^b$ ,  $W_{III}^b$  respectively. We will always hold  $b < 1$  fixed throughout, and we will see what dictates its choice later.

**Lemma 3.1.2** *If the following basic estimate on a neighbourhood  $(0 < r < b) \times N$  of the singular point (where  $N$  is the link)*

$$\|F\|_{\{\varepsilon, \varepsilon\}} \leq C\varepsilon^{1/2} (\|F\| + \|dF\|) \quad (56)$$

holds for a function  $F \in \text{dom } d$  (and  $0 < \varepsilon < b$  say) then  $\bar{d}_0 = \bar{d}$  and the laplacian  $\bar{\Delta} = \bar{\delta}^*$  is the generalised Dirichlet laplacian  $\bar{\delta}^* \bar{d}_0$ , and is therefore self-adjoint.

*Proof.* Here we have defined

$$\|F\|_{\{\eta, \varepsilon\}} = \int_N F(\varepsilon) \wedge *_\eta F(\varepsilon) dV_N$$

in the notation of [6] Lemma 5.2. For the justification of this, see [2]. Lemma 1.1. In [6], the estimate

$$\|F\|_{\{\varepsilon, \varepsilon\}} \leq C\varepsilon^{1/2}(\|F\| + \|dF\|), F \in \text{dom } d$$

is proved for algebraic surfaces. As is clear from the equations (5.5) there, it is enough to prove the estimate (56) above to conclude that there exists a sequence  $\varepsilon_n \rightarrow 0$  such that

$$\lim_{n \rightarrow \infty} \int_{r^{-1}(\varepsilon_n)} F \wedge *G = 0 \text{ for } F \in \text{dom } d \text{ and } G \in \text{dom } \delta$$

because of (5.6) there, which is also true in our case by [2], Lemma 1.2.

So we now proceed to establish the estimate (56) above. It is again enough to do for each of the  $W_I^b, W_{II}^b, W_{III}^b$  sets described above because of [6] (Lemma 5.3). In accordance with the lemma 2.3.2, since the estimate (56) we seek depends only on the quasimetric class of the metric, and since by inequality (48) of the same lemma we have  $0 < r < b \Leftrightarrow 0 < r_1 < b'$  for some  $b'$  depending on  $b$ , we may as well work with the local coordinate  $r_1 = |\zeta_1|$  instead of the global distance coordinate  $r$ , and redefine  $W_\alpha^b$  for  $\alpha = I, II, III$  as  $W \cap (0 < r_1 < b)$  where  $W$  is the small  $(u, v, w)$  polydisks of 2.3.1 and 2.3.2.

### PROPOSITION 3.1.3

*The basic estimate (56) is valid on sets of the type  $W_I^b$ .*

*Proof.* Referring to 2.3.1 above  $\zeta_1 = u^{a_1}, \zeta_2 = \zeta_1^\alpha v, \zeta_3 = \zeta_1^\beta w$ , where  $\alpha = \frac{a_2}{a_1}$  and  $\beta = \frac{a_3}{a_1}$

$\theta_1 = \arg \zeta_1$ . Also denote  $x_i = \text{Re } \zeta_i$  and  $y_i = \text{Im } \zeta_i$  for  $i = 2, 3$ .

The metric on

$$W_I^b = (0 < r_1 < b) \times S^1 \times (v: 0 \leq |v| < 1) \times (w: 0 \leq |w| < 1)$$

is quasi-isometric to (by 2.3.1)

$$dr^2 + r_1^2 d\theta_1^2 + r_1^{2\alpha} (dx_1^2 + dy_1^2) + r_1^{2\beta} (dx_2^2 + dy_2^2) \quad (5)$$

This follows from (9) and 2.3.1 by straightforward substitutions of  $\zeta_i$  in terms of the variables defined above, and bounding of cross terms by the Cauchy-Schwarz inequality as in Lemma 3.2 of [5]. (Of course, one may have to choose  $b$  to be small enough to do this).



For notational ease, we will drop the subscript on  $r_1$  from now on. Note that  $r^{-1}(b)$  is a copy of the link part, viz.  $N \cap W_r^b$ .

Getting to the proof of (56) on  $W_r^b$ , we have the volume form, from (57) above as

$$d \text{Vol} = r^{2\alpha+2\beta+1} dr d\theta_1 dx_1 dy_1 dx_2 dy_2 = r^{2\alpha+2\beta+1} dV_N$$

As in [6] 5.4, we define for a function  $F$ ,

$$\begin{aligned} \|F\|_{\{\eta, \varepsilon\}}^2 &= \int_N |F(\varepsilon, \theta_1, x_1 y_1, x_2, y_2)|^2 \eta^{2\alpha+2\beta+1} dV_N \\ &= C \eta^{2\alpha+2\beta+1} \|F\|_{\{b, \varepsilon\}}^2 \end{aligned} \quad (58)$$

where  $C = b^{-2\alpha-2\beta-1}$  is a fixed constant depending only on  $b$ .

Similarly for a 1-form  $w dr$ ,

$$\begin{aligned} \|w dr\|_{\{\eta, \varepsilon\}}^2 &= C \eta^{2\alpha+2\beta+1} \|w dr\|_{\{b, \varepsilon\}}^2 \\ &= C \eta^{2\alpha+2\beta+1} \|w\|_{\{b, \varepsilon\}}^2 \end{aligned} \quad (59)$$

where  $\|w\|_{\{b, \varepsilon\}}^2$  is defined above in (58). In the sequel,  $C$  will denote a generic constant.

Now for  $0 < \varepsilon < a \leq b$ , and  $F \in \text{dom } d$ , we have from (58) and (59)

$$\begin{aligned} \left\| \int_{\varepsilon}^a \frac{\partial F}{\partial r} dr \right\|_{\{\varepsilon, \varepsilon\}} &= C^{1/2} \varepsilon^{\alpha+\beta+1/2} \left\| \int_{\varepsilon}^a \frac{\partial F}{\partial r} dr \right\|_{\{b, \varepsilon\}} \\ &\leq C^{1/2} \varepsilon^{\alpha+\beta+1/2} \int_{\varepsilon}^a \left\| \frac{\partial F}{\partial r} \right\|_{\{b, r\}} dr \\ &\leq C^{1/2} \varepsilon^{\alpha+\beta+1/2} \int_{\varepsilon}^a C^{-1/2} r^{-\alpha-\beta-1/2} \left\| \frac{\partial F}{\partial r} \right\|_{\{r, r\}} dr \\ &\leq \varepsilon^{\alpha+\beta+1/2} \left[ \int_{\varepsilon}^a r^{-2\alpha-2\beta-1} dr \right]^{1/2} \left[ \int_{\varepsilon}^a \left\| \frac{\partial F}{\partial r} \right\|_{\{r, r\}}^2 dr \right]^{1/2} \\ &\leq C \varepsilon^{\alpha+\beta+1/2} \varepsilon^{-\alpha-\beta} \left[ 1 - \left( \frac{\varepsilon}{a} \right)^{2\alpha+2\beta} \right]^{1/2} \|dF\| \\ &\leq C \varepsilon^{1/2} \|dF\|. \end{aligned} \quad (60)$$

If  $a \in \left[ \frac{b}{2}, b \right]$  is the point where the function (of  $r$ )  $\|F\|_{\{b, r\}}$  take its minimum we have by (58),

$$\begin{aligned} \|F(a)\|_{\{\varepsilon, \varepsilon\}} &= \|F\|_{\{\varepsilon, a\}} \leq C \varepsilon^{\alpha+\beta+1/2} \|F\|_{\{b, a\}} \\ &\leq C \varepsilon^{\alpha+\beta+1/2} \int_{b/2}^b \|F\|_{\{b, r\}} dr \\ &\leq C \varepsilon^{\alpha+\beta+1/2} \left[ \int_{b/2}^b r^{-2\alpha-2\beta-1} dr \right]^{1/2} \left[ \int_{b/2}^b \|F\|_{\{r, r\}}^2 dr \right]^{1/2} \\ &\leq C \varepsilon^{1/2} \|F\| \end{aligned} \quad (61)$$

since  $\alpha, \beta \geq 1$ .

Since

$$\|F\|_{\{\varepsilon, \varepsilon\}} \leq \left\| \int_{\varepsilon}^a \frac{\partial F}{\partial r} dr \right\|_{\{\varepsilon, \varepsilon\}} + \|F(a)\|_{\{\varepsilon, \varepsilon\}}$$

the required inequality (56) is established.  $\square$

We now deal with double point charts.

#### PROPOSITION 3.1.4

The basic estimate (56) is valid for regions of the type  $W_{II}^b$ .

*Proof.* We first construct a coordinate system which expresses

$$W_{II}^b = (0 < r_1 < b) \times N$$

as a product. Going back to the description of 2.3.1 above, we may assume that  $a_1 b_2 - a_2 b_1$  is positive (by interchanging  $u$  and  $v$  necessary). Recall that

$$\zeta_1 = u^{a_1} v^{b_1}, \zeta_2 = u^{a_2} v^{b_2}, \zeta_3 = u^{a_3} v^{b_3} w$$

where  $a_3 \geq a_2 \geq a_1 \geq 1$ ,  $b_3 \geq b_2 \geq b_1 \geq 1$ . We relabel  $|\zeta_1| = r_1$  and  $r$  as in the last proposition for notational convenience. Hence if we denote

$$\alpha_1 = \frac{a_2}{a_1} \geq 1, \beta_1 = \frac{b_2}{b_1} \geq 1, \alpha_2 = \frac{a_3}{a_1} \geq \alpha_1 \geq 1, \text{ and } \beta_2 = \frac{b_3}{b_1} \geq \beta_1 \geq 1$$

we have by the assumption above that  $\alpha_1 < \beta_1$ .

Consider the coordinate system  $(r, s, \sigma, \theta_1, \theta_2, \theta_3)$  where  $r, \sigma, \theta_i$  are defined in 3.1.1 and  $s$  is defined by

$$s = \frac{a_2 |\log \rho| + b_2 |\log \tau|}{a_1 |\log \rho| + b_1 |\log \tau|} \quad (62)$$

which is well defined for  $0 < \rho < 1$ ,  $0 < \tau < 1$ , so on  $W_{II}^b$ . Also clearly  $\alpha_1 < s < \beta_1$ , because for any fixed  $(r, \sigma, \theta_i)$ , by making

$$\frac{\log \tau}{\log \rho} \left( \text{resp. } \frac{\log \rho}{\log \tau} \right)$$

as small as we want, we can make  $s$  as close to  $\alpha_1$  (resp.  $\beta_1$ ) as we want. See the figure 1 below for a picture of the  $\rho, \tau$  cross-section of  $W_{II}^b$ . Thus  $r_2 = |\zeta_2| = r_1^s \stackrel{\text{def}}{=} r^s$ . Further since by assumption the determinant  $a_1 b_2 - a_2 b_1 > 0$ , we have

$$(a_3, b_3) = \lambda_1 (a_1, b_1) + \lambda_2 (a_2, b_2)$$

so that

$$r_3 = |\zeta_3| = r^{\lambda_1} r_2^{\lambda_2} \sigma \text{ where } r^{\lambda_1} r_2^{\lambda_2} = \rho^{a_3} \tau^{b_3} < 1$$

This implies

$$dr_2 = sr^{s-1} dr + r^s |\log r| ds$$

$$dr_3 = \lambda_1 \frac{r_3}{r} dr + \lambda_2 \frac{r_3}{r_2} dr_2 + r^{\lambda_1} r_2^{\lambda_2} d\sigma \quad (63)$$

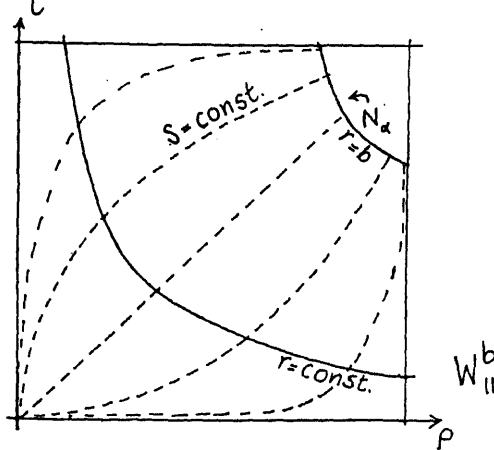


Figure 1.

Thus the original metric  $\Sigma_{i=1}^3 d\zeta_i d\bar{\zeta}_i$  of (2.3.1) which is

$$\sum_{i=1}^3 (dr_i^2 + r_i^2 d\theta_i^2)$$

is quasi-isometric to (using (63))

$$dr^2 + r^2 d\theta_1^2 + dr_2^2 + r_2^2 d\theta_2^2 + r^{2\lambda_1} r_2^{2\lambda_2} (d\sigma^2 + \sigma^2 d\theta_3^2) \quad (64)$$

by using the Cauchy-Schwartz inequality to absorb all  $dr dr_2$ ,  $dr d\sigma$ , and  $dr_2 d\sigma$  cross terms by noting that  $\frac{r_3}{r}$  and  $\frac{r_2}{r}$  can be made arbitrarily small by making  $\sigma$  range in a small enough interval (see [5], §3). Now using the first equation of (63), and the fact that  $r^{s-1}$  is bounded since  $s > \alpha_1 \geq 1$ , and again invoking the Cauchy-Schwartz inequality to remove cross terms involving  $ds dr$ , we plug  $r_2 = r^s$  in (64) above to get

$$W_{II}^b = (0 < r < b) \times (\alpha_1 < s < \beta_1) \times (0 \leq \sigma < A) \times \{(\theta_1, \theta_2, \theta_3) \in T^3\}$$

with the metric quasi-isometric to

$$dr^2 + r^2 d\theta_1^2 + r^{2s} ((\log r)^2 ds^2 + d\theta_2^2) + r^{2(\lambda_1 + s\lambda_2)} (d\sigma^2 + \sigma^2 d\theta_3^2) \quad (65)$$

where  $\lambda_1 + s\lambda_2$  lies in the range  $(\alpha_2, \beta_2)$ , noting that  $\alpha_2 = \frac{a_3}{a_1} = \lambda_1 + \alpha_1 \lambda_2 \geq 1$  and  $\beta_2 = \frac{b_3}{b_1} = \lambda_1 + \alpha_2 \lambda_2 \geq 1$ . (To be specific, we have assumed  $\beta_2 > \alpha_2$ ; there is no change in the arguments if the reverse inequality is assumed).

*Remark 3.1.5* In the above expression (65), if we put  $\sigma = \theta_3 = 0$ , we will have an alternative metric description of a neighborhood of a normal crossing in the surface case, referred to as  $W(+)$  in [6], to the one given there.

*Lemma 3.1.6.* Let  $0 < \varepsilon < a < c = \frac{1}{\varepsilon}$ . Then for any constant  $\mu \geq 5$ , we have the integral estimates

$$\int_{\varepsilon}^a r^{-\mu} |\log r|^{-1} dr \leq K \varepsilon^{1-\mu} |\log \varepsilon|^{-1} \quad (66)$$

$$\int_{\varepsilon}^a r^{-\mu} |\log r|^{-2} dr \leq K \varepsilon^{1-\mu} |\log \varepsilon|^{-2} \quad (67)$$

where  $K$  is a constant independent of  $\mu$ .

*Proof.* Since the integrands are positive, enough to replace  $a$  by  $c = \frac{1}{\varepsilon}$ . Integrating by parts, noting

$$\frac{d}{dr} |\log r|^{-1} = r^{-1} |\log r|^{-2}$$

we get

$$\int_{\varepsilon}^c r^{-\mu} |\log r|^{-1} dr = \left| |\log r|^{-1} \frac{r^{1-\mu}}{1-\mu} \right|_{\varepsilon}^c + \frac{1}{\mu-1} \int_{\varepsilon}^c r^{-\mu} |\log r|^{-2} dr$$

So that

$$\int_{\varepsilon}^c r^{-\mu} |\log r|^{-1} \left( 1 - \frac{|\log r|^{-1}}{\mu-1} \right) dr = \varepsilon^{1-\mu} \frac{|\log \varepsilon|^{-1}}{\mu-1} \left( 1 - \left( \frac{\varepsilon}{c} \right)^{\mu-1} |\log \varepsilon| \right).$$

Now since  $\mu \geq 5$ , we have  $\frac{1}{\mu-1} \leq \frac{1}{4}$ , and

$$c = \frac{1}{\varepsilon} \Rightarrow \left( 1 - \frac{|\log r|^{-1}}{\mu-1} \right) \geq \left( 1 - \frac{|\log c|^{-1}}{4} \right) \geq \frac{3}{4}$$

for  $r \in (\varepsilon, c)$  and  $\mu \geq 5$ .

Also

$$0 < 1 - \left( \frac{\varepsilon}{c} \right)^{\mu-1} |\log \varepsilon| < 1$$

since  $x^{\mu-1} |\log x|$  is an increasing function on  $(0, c)$  which implies that

$$0 < \frac{\varepsilon^{\mu-1} |\log \varepsilon|}{c^{\mu-1}} = \frac{\varepsilon^{\mu-1} |\log \varepsilon|}{c^{\mu-1} |\log c|} < 1$$

This proves (66). The proof of (67) is similar, and is omitted. □

We now resume the proof of Proposition 3.1.4. By (65), the volume form is given by:

$$d \text{Vol} = r^{2s+1+2(s\lambda_2+\lambda_1)} |\log r| d\theta_1 d\theta_2 d\theta_3 \sigma d\sigma dr ds$$

We abbreviate  $\sigma d\sigma d\theta_2 d\theta_3$  as  $dk$ . Since  $b$  is fixed, and for  $s \in (\alpha_1, \beta_1)$ , the quantity  $(s\lambda_2 + \lambda_1) \in (\alpha_2, \beta_2)$  as pointed out in (65),  $b^{2s+1+2(s\lambda_2+\lambda_1)} |\log b|$  is bounded on both sides by positive constants  $C'_1$  and  $C'_2$  for all  $s \in (\alpha_1, \beta_1)$ . Hence defining

$$\|f\|_{\{r,\eta,s\}}^2 = r^{2s+1+2(s\lambda_2+\lambda_1)} |\log r| \int_{\theta_i, \sigma} |f(\eta, s, \sigma, \theta_i)|^2 dk$$

we have the inequalities, for all  $s \in (\alpha_1, \beta_1)$ .

$$\begin{aligned} C_2 r^{s+1/2+(s\lambda_2+\lambda_1)} |\log r|^{1/2} \|f\|_{\{b,\eta,s\}} &\leq \|f\|_{\{r,\eta,s\}} \\ &\leq C_1 r^{s+1/2+(s\lambda_2+\lambda_1)} |\log r|^{1/2} \|f\|_{\{b,\eta,s\}} \end{aligned} \quad (68)$$

where  $C_i$  are some positive constants (in fact  $C_i = (C'_i)^{-1/2}$ ). Note also that

$$\begin{aligned} \|f\|_{\{r,\eta\}}^2 &= \int_{\alpha_1}^{\beta_1} \|f\|_{\{r,\eta,s\}}^2 ds \\ \|f\|^2 &= \int_0^b \|f\|_{\{r,r\}}^2 dr \end{aligned} \quad (69)$$

With corresponding obvious definitions, the inequalities (68), and (69) are also checked to apply to a 1-form of the type  $w dr$ , which we shall designate as (68b) and (69b) respectively.

Let  $0 < \varepsilon < a < b \leq c = \frac{1}{e}$ . Then by (68), (68b), we have

$$\begin{aligned} \left\| \int_{\varepsilon}^a \frac{\partial F}{\partial r} dr \right\|_{\{e,e,\bar{s}\}} &\leq C_1 \varepsilon^{s+1/2+(s\lambda_2+\lambda_1)} |\log \varepsilon|^{1/2} \left\| \int_{\varepsilon}^a \frac{\partial F}{\partial r} dr \right\|_{\{b,e,s\}} \\ &\leq C_1 \varepsilon^{s+1/2+(s\lambda_2+\lambda_1)} |\log \varepsilon|^{1/2} \int_{\varepsilon}^a \left\| \frac{\partial F}{\partial r} \right\|_{\{b,r,s\}} dr \\ &\leq \frac{C_1}{C_2} \varepsilon^{s+1/2+(s\lambda_2+\lambda_1)} |\log \varepsilon|^{1/2} \int_{\varepsilon}^a r^{-s-1/2+(s\lambda_2+\lambda_1)} |\log r|^{-1/2} \left\| \frac{\partial F}{\partial r} \right\|_{\{r,r,s\}} dr \end{aligned}$$

which, by the Cauchy-Schwartz inequality is

$$\begin{aligned} &\leq C \varepsilon^{s+1/2+(s\lambda_2+\lambda_1)} |\log \varepsilon|^{1/2} \left( \int_{\varepsilon}^a r^{-2s-(s\lambda_2+\lambda_1)} |\log r|^{-1} \right)^{1/2} \\ &\quad \times \left( \int_{\varepsilon}^a \left\| \frac{\partial F}{\partial r} \right\|_{\{r,r,s\}}^2 dr \right)^{1/2} \end{aligned} \quad (70)$$

Now  $(s\lambda_2 + \lambda_1) \geq \alpha_2 \geq 1$  and  $s \geq \alpha_1 \geq 1$ . Thus

$$\mu = 2s + 1 + 2(s\lambda_2 + \lambda_1) \geq 5$$

By Lemma 3.1.6 above since  $b \leq c \stackrel{\text{def}}{=} e^{-1}$ , and (66), the expression in (70), for all

$$\leq C\varepsilon^{s+1/2+(s\lambda_2+\lambda_1)}|\log \varepsilon|^{1/2}\varepsilon^{-s-(s\lambda_2+\lambda_1)}|\log \varepsilon|^{-1/2}\left(\int_0^b\left\|\frac{\partial F}{\partial r}\right\|_{\{r,r,s\}}^2dr\right)^{1/2}$$

$$\leq C\varepsilon^{1/2}\left(\int_0^b\left\|\frac{\partial F}{\partial r}\right\|_{\{r,r,s\}}^2dr\right)^{1/2}$$

Thus

$$\left\|\int_\varepsilon^a\frac{\partial F}{\partial r}dr\right\|_{\{r,r,s\}}^2\leq C\varepsilon\int_0^b\left\|\frac{\partial F}{\partial r}\right\|_{\{r,r,s\}}^2dr. \quad (71)$$

Now, for a fixed  $s\in(\alpha_1,\beta_1)$ , let  $a_s$  be the point where  $\|F\|_{\{b,r,s\}}$  (as a function of  $r$ ) achieves its minimum on  $\left[\frac{b}{2},\frac{2b}{3}\right]$ . Then by (68)

$$\begin{aligned}\|F(a_s)\|_{\{\varepsilon,e,s\}} &\stackrel{\text{def}}{=} \|F\|_{\{\varepsilon,a_s,s\}} \\ &\leq C_1\varepsilon^{s+1/2+(s\lambda_2+\lambda_1)}|\log \varepsilon|^{1/2}\|F\|_{\{b,a_s,s\}} \\ &\leq \frac{6C_1}{b}\varepsilon^{s+1/2+(s\lambda_2+\lambda_1)}|\log \varepsilon|^{1/2}\int_{b/2}^{2b/3}\|F\|_{\{b,r,s\}}dr\end{aligned} \quad (72)$$

$$\begin{aligned}&\leq C'\varepsilon^{s+1/2+(s\lambda_2+\lambda_1)}|\log \varepsilon|^{1/2}\left(\int_{b/2}^{2b/3}r^{-2s-1-2(s\lambda_2+\lambda_1)}|\log r|^{-1}dr\right)^{1/2} \\ &\quad\times\left(\int_0^b\|F\|_{\{r,r,s\}}^2dr\right)^{1/2}\end{aligned} \quad (73)$$

by (68) again, and the Cauchy-Schwartz inequality. Now, noting that for all  $s\in(\alpha_1,\beta_1)$ ,  $2s+1+2(s\lambda_2+\lambda_1)\geq 5$ , the integral in the first parenthesis of (73) is  $\leq C$  (independent of  $s$ ). Thus

$$\|F(a_s)\|_{\{\varepsilon,e,s\}}^2\leq C\varepsilon^{s+1/2+(s\lambda_2+\lambda_1)}|\log \varepsilon|^{1/2}\int_0^b\|F\|_{\{r,r,s\}}^2dr \quad (74)$$

But again

$$\varepsilon^{2s+2(s\lambda_2+\lambda_1)}|\log \varepsilon|^{1/2}\leq \varepsilon^4|\log \varepsilon|^{1/2}$$

by (71) above, and for  $\varepsilon\in(0,1)$ , this is less than equal to some constant  $C'$ , we have

$$\|F(a_s)\|_{\{\varepsilon,e,s\}}^2\leq C\varepsilon\int_0^b\|F\|_{\{r,r,s\}}^2dr \quad (75)$$

Finally,

$$\begin{aligned}\|F\|_{\{e,e\}}^2 &= \int_{\alpha_1}^{\beta_1}\|F\|_{\{e,e,s\}}^2ds \\ &\leq 2\int_{\alpha_1}^{\beta_1}\left(\left\|\int_\varepsilon^s\frac{\partial F}{\partial r}dr\right\|_{\{e,e,s\}}^2+\|F(a_s)\|_{\{e,e,s\}}^2\right)ds \\ &\leq C\varepsilon\int_{\alpha_1}^{\beta_1}\left(\int_0^b\left\|\frac{\partial F}{\partial r}\right\|_{\{r,r,s\}}^2dr+\int_0^b\|F\|_{\{r,r,s\}}^2dr\right)ds\end{aligned}$$

by the Cauchy-Schwartz inequality and (71), (75) above. Thus

$$\begin{aligned}
 \|F\|_{\{\varepsilon, \varepsilon\}}^2 &\leq C\varepsilon \int_0^b \left( \int_{\alpha_1}^{\beta_1} \left\| \frac{\partial F}{\partial r} \right\|_{\{r, r, s\}}^2 ds + \int_{\alpha_1}^{\beta_1} \|F\|_{\{r, r, s\}}^2 ds \right) dr \\
 &= C\varepsilon \int_0^b \left( \left\| \frac{\partial F}{\partial r} \right\|_{\{r, r\}}^2 + \|F\|_{\{r, r\}}^2 \right) dr \\
 &\leq C\varepsilon (\|dF\|^2 + \|F\|^2) \\
 &\leq C\varepsilon (\|dF\| + \|F\|)^2
 \end{aligned} \tag{76}$$

by (69), which implies (3.11) as claimed.  $\square$

*Remark 3.1.7.* We remark that in view of remark 3.1.5 above, the proof above for 3.1.4 will apply to sets of the type  $W(+)$  at a normal crossing for the surface case. (See [6] for another proof.) All we would need to do is set  $\theta_3 = \sigma = 0$  and improve the Lemma 3.1.6 for  $\mu \geq 3$ , by taking a different choice of  $b$ .

### PROPOSITION 3.1.8

The basic estimate (56) is true for the sets  $W_{III}^b$ .

*Proof.* First we give the product decomposition of  $W_{III}^b$ . Consider the new variables

$$s = \frac{a_2 |\log \rho| + b_2 |\log \tau| + c_2 |\log \sigma|}{a_1 |\log \rho| + b_1 |\log \tau| + c_1 |\log \sigma|}$$

and

$$t = \frac{a_3 |\log \rho| + b_3 |\log \tau| + c_3 |\log \sigma|}{a_1 |\log \rho| + b_1 |\log \tau| + c_1 |\log \sigma|}$$

which make sense in  $0 < \rho < 1$ ,  $0 < \tau < 1$ ,  $0 < \sigma < 1$  and hence on  $W_{III}^b$ . Here  $r = r_1 = \rho^{a_1} \tau^{b_1} \sigma^{c_1}$ . If we denote by

$$\begin{bmatrix} \lambda_1 & \lambda_2 & \lambda_3 \\ \mu_1 & \mu_2 & \mu_3 \\ \nu_1 & \nu_2 & \nu_3 \end{bmatrix}$$

the inverse of matrix

$$\begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix}$$

we see that

$$|\log \rho| = |\log r|(\lambda_1 + \lambda_2 s + \lambda_3 t)$$

$$|\log \tau| = |\log r|(\mu_1 + \mu_2 s + \mu_3 t)$$

$$|\log \sigma| = |\log r|(\nu_1 + \nu_2 s + \nu_3 t)$$

Since  $|\log \rho|, |\log \tau|, |\log \sigma| > 0$ , we see that  $\lambda_1 + \lambda_2 s + \lambda_3 t > 0$ ,  $\mu_1 + \mu_2 s + \mu_3 t > 0$ ,

sides are given by the equations  $\lambda_1 + \lambda_2 s + \lambda_3 t = 0$ ,  $\mu_1 + \mu_2 s + \mu_3 t = 0$ , and  $v_1 + v_2 s + v_3 t = 0$ . Further, for every  $(s, t) \in A$ , and every  $r \in (0, b)$  we may use the above for  $|\log \rho|$ ,  $|\log \tau|$ ,  $|\log \sigma|$ . This solution for  $(\rho, \tau, \sigma)$  satisfies  $r = \rho^{a_1} \tau^{b_1} \sigma^{c_1}$  as is seen by multiplying the three equations above by  $a_1$ ,  $b_1$ ,  $c_1$  respectively and then adding. Note the vertices of  $A$  are

$$\begin{pmatrix} \alpha_1 = \frac{a_2}{a_1}, & \alpha_2 = \frac{a_3}{a_1} \\ \beta_1 = \frac{b_2}{b_1}, & \beta_2 = \frac{b_3}{b_1} \\ \gamma_1 = \frac{c_2}{c_1}, & \gamma_2 = \frac{c_3}{c_1} \end{pmatrix}$$

Thus

$$W_{III}^b = (0 < r < b) \times ((s, t) \in A) \times ((\theta_1, \theta_2, \theta_3) \in T^3)$$

as a product. Again  $r_2 = r_1^s = r^s$ ,  $r_3 = r^t$ . Now we may proceed exactly as in the proof of (65) to make the metric

$$\sum_{i=1}^3 d\zeta_i d\bar{\zeta}_i = \sum_{i=1}^3 (dr_i^2 + r_i^2 d\theta_i^2)$$

quasi-isometric to

$$\begin{aligned} dr^2 + r^2 d\theta_1^2 + r^{2s}(|\log r|^2 ds^2 + d\theta_2^2) + r^{2t}(|\log r|^2 dt^2 + d\theta_3^2) \\ \Rightarrow dVol = r^{2s+2t+1} |\log r|^2 dr ds dt d\theta_1 d\theta_2 d\theta_3 \end{aligned} \quad (77)$$

We can now prove 3.1.8. For  $(s, t) \in A$ , define

$$\|F\|_{\{r, \eta, s, t\}}^2 = r^{2s+2t+1} |\log r|^2 \int_{\theta_i} |F(\eta, s, t, \theta_i)|^2 d\theta_1 d\theta_2 d\theta_3$$

and

$$\|F\|_{\{r, \eta\}}^2 = \int_{(s, t) \in A} \|F\|_{\{r, \eta, s, t\}}^2 ds dt$$

So that

$$\|F\|^2 = \int_0^b \|F\|_{\{r, \eta\}}^2 dr$$

Now since  $(s, t) \in A$  whose vertices are  $(\alpha_1, \alpha_2)$ ,  $(\beta_1, \beta_2)$ ,  $(\gamma_1, \gamma_2)$ , where all the  $\alpha_i$ 's,  $\beta_i$ 's,  $\gamma_i$ 's are  $\geq 1$ , we see that

$$2s + 2t + 1 \geq 5$$

for  $(s, t) \in A$ . Also

$$2s + 2t + 1 \leq 2\alpha + 2\beta + 1$$

where  $\alpha = \max(\alpha_1, \beta_1, \gamma_1)$ ,  $\beta = \max(\alpha_2, \beta_2, \gamma_2)$ .

In complete analogy with the proof of (71) (and using (67) instead of (66)) we get



for  $0 < \varepsilon < a < b < 1$

$$\left\| \int_{\varepsilon}^a \frac{\partial F}{\partial r} dr \right\|_{\{e, e, s, t\}}^2 \leq C\varepsilon \int_0^b \left\| \frac{\partial F}{\partial r} \right\|_{\{r, r, s, t\}}^2 dr \quad (78)$$

Similarly, in analogy with the proof of (75), we get

$$\|F(a_{s,t}\|_{\{e, e, s, t\}}^2 \leq C\varepsilon \int_0^b \|F\|_{\{r, r, s, t\}}^2 dt \quad (79)$$

where  $a_{s,t}$  is the point in  $\left[\frac{b}{2}, \frac{2b}{3}\right]$  where  $\|F\|_{\{b, r, s, t\}}^2$  reaches its minimum as a function of  $r$ .

Finally

$$\|F\|_{\{e, e\}}^2 = \int_{(s,t) \in A} \|F\|_{\{e, e, s, t\}}^2 ds dt$$

Using (78), (79) above, and interchanging the  $ds dt$  integral over  $A$  and  $dr$  integral over  $(0, b)$ , exactly by the steps leading up to (76), we have

$$\|F\|_{\{e, e\}}^2 \leq C\varepsilon (\|dF\|^2 + \|F\|^2)$$

which proves estimate (56) and the Proposition 3.1.8. Thus the main theorem 1.1 is established.

## 4. Appendix

### 4.1 A transversality lemma for isolated threefold singularities

A J PARAMESWARAN and V SRINIVAS

School of mathematics,

Tata Institute of Fundamental Research

Bombay 400 005, India

#### PROPOSITION 4.1.1

Let  $X$  be a representative of the isolated threefold singularity germ  $(X, 0)$ . Then, after a linear change of coordinates, and possibly shrinking the representative  $x$ , the following is true: (see § 2.2 for notation)

- (i)  $D_i$  is a smooth surface for all  $1 \leq i \leq N$
- (ii)  $E_{ijk}$  is a smooth for all  $1 \leq i \leq j \leq k \leq N$
- (iii)  $D_i$  and  $D_j$  meet transversally for all  $i \neq j$ .
- (iv)  $E_{ijk}$ , and  $E_{i',j',k'}$  meet transversally for all  $(i, j, k) \neq (i', j, k')$
- (v)  $E_{ijk}$  and  $D_l$  meet transversally for all  $(i, j, k)$  and all  $l$ .

(N.B. Here, some of the coordinates  $(i, j, k)$ ,  $(i' j' k')$  in (iv) are allowed to coincide, and in (v), the cases  $l = i, j, k$  are allowed.)

We will need the following Lemma, which is a slight generalisation of Kleiman's Bertini Theorem.

**Lemma 4.1.2.** *Let  $Y$  be a complex manifold, and let  $\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_k$  be locally free sheaves of ranks  $s_1, s_2, \dots, s_k$  respectively. Let  $V_i$  be an  $n_i$ -dimensional space of global sections of  $\mathcal{E}_i$  which generates  $\mathcal{E}_i$ , for each  $1 \leq i \leq k$ . Let  $r_i$  be a positive integer, with  $r_i \leq n_i$ , and set  $m_i = \min\{r_i, s_i\}$ ; let  $G_i$  be the grassmannian of  $r_i$ -dimensional spaces of  $V_i$ , and set  $G = \prod_{i=1}^k G_i$ . For  $t = (t_1, \dots, t_k) \in G$ , if  $W_i(t) \subset V_i$  are the corresponding subspaces, let*

$$Y_t(\mathcal{E}_1, \dots, \mathcal{E}_k) = \{y \in Y \mid \text{for each } 1 \leq i \leq k, W_i(t) \otimes \mathcal{O}_{Y,y} \rightarrow \mathcal{E}_{i,y} \text{ has rank } < m_i\}$$

with its natural (possibly non-reduced) structure as an analytic space, defined locally in  $Y$  by the vanishing of determinants.

Then there is a dense subset  $U \subset G$ , whose complement is a countable union of locally closed analytic subsets of  $G$  of smaller dimension, such that for  $t \in U$ .

(i)  $Y_t(\mathcal{E}_1, \dots, \mathcal{E}_k)$  is empty, or has codimension

$$\sum_{i=1}^k (r_i - m_i + 1)(s_i - m_i + 1)$$

in  $Y$ .

(ii) the singular locus of  $Y_t(\mathcal{E}_1, \dots, \mathcal{E}_k)$  is empty, or has codimension

$$\sum_{i=1}^k (r_i - m_i + 2)(s_i - m_i + 2)$$

in  $Y$ .

(iii) the singular locus of the singular locus of  $Y_t(\mathcal{E}_1, \dots, \mathcal{E}_k)$  is empty, or has codimension

$$\sum_{i=1}^k (r_i - m_i + 3)(s_i - m_i + 3)$$

in  $Y$ .

(iv)  $Y_t(\mathcal{E}_1, \dots, \mathcal{E}_k)$  is Cohen-Macaulay (i.e., its local rings are Cohen-Macaulay).

**Proof.** Let  $G'_i$  be the grassmannian of  $s_i$ -dimensional quotients of  $V_i$ , and let  $f_i: Y \rightarrow G'_i$  be the classifying map of  $\mathcal{E}_i$ . Let

$$\Gamma_Y = \{(t, y) \in G \times Y \mid y \in Y_t(\mathcal{E}_1, \dots, \mathcal{E}_k)\}$$

If  $G' = \prod_{i=1}^k G'_i$ , and  $Q_i$  is the universal quotient bundle of rank  $s_i$  on the grassmannian  $G'_i$ , then there is a universal such subvariety

$$\Gamma = \{(t, y) \in G \times G' \mid y \in G'_i(Q_1, \dots, Q_k)\}$$

which is  $\Gamma_Y$  for the case  $Y = G'$ ,  $\mathcal{E}_i = Q_i$ .

The projections  $\Gamma \rightarrow G$ ,  $\Gamma \rightarrow G'$  are known to be locally trivial fibre bundles (for the Zariski topology), whole fibres are irreducible and Cohen-Macaulay; further,

(a)  $\Gamma$  has codimension

$$\sum_{i=1}^k (r_i - m_i + 1)(s_i - m_i + 1)$$

in  $G \times G'$ .

(b) the singular locus  $\Gamma_{\text{sing}}$  of  $\Gamma$  has codimension

$$\sum_{i=1}^k (r_i - m_i + 2)(s_i - m_i + 2)$$

in  $\mathbf{G} \times \mathbf{G}'$ , since it is the locus of pairs  $(t, y)$  where for each  $i$ , the map  $W_i(t) \rightarrow Q_i \otimes C(y_i)$  has rank  $\leq m_i - 2$ .

(c) the singular locus of  $\Gamma_{\text{sing}}$  has codimension  $\sum_{i=1}^k (r_i - m_i + 3)(s_i - m_i + 3)$  in  $\mathbf{G} \times \mathbf{G}'$ , since it is the locus of pairs  $(t, y)$  where for each  $i$  the rank of the map  $W(t) \rightarrow Q_i \otimes C(y_i)$  is  $\leq m_i - 3$ .

These assertions are easily reduced to the case  $k = 1$ , since

$$G'_t(Q_1, \dots, Q_k) = \prod_{i=1}^k (G'_i)_{t_i}(Q_i)$$

(For the case  $k = 1$ , see, for example, Fulton's *Intersection Theory*, Ch. 14, especially Th. 14.3). Hence, the same estimates hold for the codimensions in  $\mathbf{G}$  of any fibre of  $\Gamma \rightarrow \mathbf{G}'$ , and for that of its singular locus, etc.

Now  $\Gamma_Y = \Gamma \times_{\mathbf{G}} Y \rightarrow Y$  is also a locally trivial fibre bundle over the complex manifold  $Y$ , so the same estimates hold for the codimensions in  $\mathbf{G} \times Y$  of  $\Gamma_Y$ , that of its singular locus, etc. The first three assertions of the lemma now follow from Sard's theorem applied to the projection to  $\mathbf{G}$ . Since  $\Gamma, Y$  are Cohen-Macaulay, so is  $\mathbf{G} \times Y$ ; since  $\mathbf{G} \times \mathbf{G}'$  is smooth, the fibre product  $\Gamma_Y \subset \mathbf{G} \times Y$  is locally defined by the vanishing of a regular sequence, and so is Cohen-Macaulay. Hence any fibre  $Y_t(\mathcal{E})$  of  $\Gamma_Y \rightarrow \mathbf{G}$ , which has the "expected" codimension in  $Y$ , is Cohen-Macaulay (again, since  $\mathbf{G}$  is smooth, such a fibre is defined by the vanishing of a regular sequence).  $\square$

### DEFINITION 4.1.3

If  $Y$  is an irreducible analytic space, call a subset  $U \subset Y$  *big* if it is the complement of a countable union of locally closed analytic subsets of smaller dimension. We call a morphism of irreducible analytic subspaces *dominant* if the image contains a dense open subset.

Note that if  $f: Y \rightarrow Z$  is a dominant morphism of analytic spaces, and  $U \subset Y$  is an open subset, then  $f(U) \subset Z$  contains an open subset of  $Z$ .

**Lemma 4.1.4** (i) Any countable intersection of big subsets is big.

(ii) If  $f: Y' \rightarrow Y$  is a dominant morphism of irreducible analytic spaces,  $U \subset Y$  a big subset, then  $f^{-1}(U) \subset Y'$  is big.

(iii) If  $f$  is as in (ii) with irreducible fibres,  $U' \subset Y'$  a subset whose complement is the countable union of locally closed analytic subsets, such that for some big subset  $U \subset Y$ ,  $f^{-1}(y) \cap U' \subset f^{-1}(y)$  is big for all  $y \in U$ , then  $U' \subset Y'$  is big.

More generally, suppose  $Y_1, \dots, Y_n$  are irreducible analytic spaces and  $f_i: Y_i \rightarrow Y_{i+1}$  dominant morphisms with irreducible fibres,  $U_1 \subset Y_1$  a subset whose complement is a countable union of locally closed analytic subsets. Suppose there exist subsets  $U_i \subset Y_i$  such that (i)  $U_n \subset Y_n$  is big, and (ii) for each  $2 \leq i \leq n$  and each  $x_i \in U_i$ , the subset

$$f_{i-1}^{-1}(x_i) \cap U_{i-1} \subset f_{i-1}^{-1}(x_i)$$

is big. Then  $U_1 \subset Y_1$  is big.

closed analytic subsets of  $V$ , then  $U$  fails to be big precisely when  $V - U$  contains an open subset of  $V$ . But for any dominant morphism  $f: V \rightarrow W$  of irreducible analytical spaces, if  $U$  is a subset of  $V$  whose complement has non-empty interior, then there is an open subset of  $W$  such that for all  $w$  in this set,  $f^{-1}(w) \cap U \subset f^{-1}(w)$  is not big. This easily yields (iii), and the last statement follows by induction.  $\square$

Let  $G = GL_N(\mathbf{C})$ , which we regard as parametrizing possible bases in  $\mathbf{C}^N$ . By the above remarks, it is enough to show that each of the smoothness/transversality conditions is (individually) valid for all bases of  $\mathbf{C}^N$  in some big subset of  $G$ .

We now make the following observations which, taken together, imply the proposition. Note that below, when we assert that a variety is smooth, the variety always comes equipped with a natural structure as a possibly non-reduced analytic space, and we are asserting that this analytic space is a complex manifold (in particular, it is reduced). An intersection  $S = X_1 \cap X_2$  of two analytic subsets of an irreducible analytic space  $X$  is *transversal* if  $S$  is smooth,  $X_1, X_2$  are equidimensional and smooth along  $S$ , and  $\text{codim}_X S = \text{codim}_X X_1 + \text{codim}_X X_2$  (the codimensions are measured in a neighborhood of  $S$ , and the intersection  $X_1 \cap X_2$  is defined as the analytic space whose ideal sheaf in  $X$  is  $\mathcal{I}_{X_1} + \mathcal{I}_{X_2}$  where  $\mathcal{I}_{X_i}$  is the ideal sheaf of  $X_i$ ).

1. For a given index  $i$ , there is a big subset of  $G$  on which the  $i$ -th hyperplane section of  $X - \{0\}$  is a smooth surface. To see this, apply Lemma 4.1.2 with  $Y = X - \{0\}$ ,  $k = 1$ ,  $\mathcal{E} = \mathcal{O}_{X - \{0\}}$ ,  $V = \mathbf{C}^N$  and  $r = 1$ . Here elements of  $V$  are considered as homogeneous linear functions on  $X - \{0\}$ . This gives a big subset  $U \subset \mathbf{P}^{N-1}$ , the projective space of lines in  $\mathbf{C}^N$ , parametrizing smooth hyperplane sections. The choice of the index  $i$  yields a surjective map  $\phi_i: G \rightarrow \mathbf{P}^{N-1}$ , giving it the structure of a homogeneous space; the desired big subset of  $G$  is  $\phi_i^{-1}(U)$ .
2. For given  $i \neq j$ , there is a big subset of  $G$  on which the intersection of the  $i$ -th and  $j$ -th hyperplane sections is transversal. We simply apply the Lemma 4.1.2 with  $Y = X - \{0\}$ ,  $k = 1$ ,  $\mathcal{E} = \mathcal{O}_{X - \{0\}}^{\oplus 2}$ ,  $V = \mathbf{C}^N \oplus \mathbf{C}^N$  and  $r = 1$ . Elements of  $V$  are ordered pairs of homogeneous linear functions. This yields a big subset  $U'$  of  $\mathbf{P}^{2N-1} = \mathbf{P}(\mathbf{C}^N \oplus \mathbf{C}^N)$ . The choice of indices  $i \neq j$  yields a map  $\phi_{ij}: G \rightarrow \mathbf{P}^{2N-1}$  whose image is a Zariski open subset, and  $\phi_{ij}^{-1}(U')$  is the desired big subset of  $G$ .
3. For a given triplet  $(i, j, k)$ , there is a big subset of  $G$  such that the corresponding  $E_{ijk}$  is a smooth surface.

We apply the Lemma 4.1.2 with  $Y = X - \{0\}$ ,  $k = 1$ ,  $\mathcal{E} = \Omega_{X - \{0\}}^1$ ,  $V = \mathbf{C}^N$  and  $r = 3$ . Elements of  $V$  yield differentials of homogeneous linear functions on  $X - \{0\}$ . The grassmannian  $G(3, N)$  of 3-dim. subspaces of  $V$  is homogeneous space for  $G$ , and the choice of indices yields a surjection  $\phi_{ijk}: G \rightarrow G(3, N)$ ; the inverse image of the big set of 4.1.2 under this map is the desired big subset of  $G$ .

4. For a given triplet  $(i, j, k)$  and a given index  $l \notin \{i, j, k\}$ , there is a big open subset of  $G$  such that the corresponding intersection  $E_{ijk} \cap D_l$  is transversal.  
Apply the Lemma 4.1.2 with  $Y = X - \{0\}$ ,  $k = 2$ ,  $\mathcal{E}_1 = \mathcal{O}_{X - \{0\}}$ ,  $\mathcal{E}_2 = \Omega_{X - \{0\}}^1$ ,  $V_1 \cong V_2 \mathbf{C}^N$ ,  $r_1 = 1$ ,  $r_2 = 3$ , where  $V_1$  is the space of homogeneous linear functions and  $V_2$  the space of differentials of homogeneous linear functions. Now use the fact that the choice of indices yields a map  $\psi_{l,ijk}: G \rightarrow \mathbf{P}^{N-1} \times G(3, N)$ ; since  $l \notin \{i, j, k\}$ , its image is a Zariski open subset.

5. For a given pair of triplets  $(i, j, k)$  and  $(i', j', k')$  with no common indices, there is a big subset of  $G$  on which the intersection  $E_{ijk} \cap E_{i'j'k'}$  is transversal. We apply the Lemma 4.1.2 with  $Y = X - \{0\}$ ,  $k = 2$ ,  $\mathcal{E}_1 \cong \mathcal{E}_2 \cong \Omega_{X-\{0\}}^1$ ,  $V_1 \cong V_2 \cong \mathbb{C}^N$  (considered as differentials of homogeneous linear functions),  $r_1 = r_2 = 3$ . The choice of indices yields a map  $\psi_{ijk, i'j'k'}: G \rightarrow G(3, N) \times G(3, N)$  whose image is a Zariski open subset, since all six indices are distinct.
6. Let  $i, j, k$  be distinct indices, and choose  $l \in \{i, j, k\}$ ; then there is a big subset of  $G$  for which  $D_l$  intersects  $E_{ijk}$  transversally outside a finite set. The choice of indices yields a map  $\psi_{l,ijk}: G \rightarrow \mathbb{P}^{N-1} \times G(3, N)$  whose image is a closed (flag) subvariety  $F(1, 3)$  consisting of pairs of subspaces  $(L, W)$  of  $\mathbb{C}^N$  with  $\dim L = 1$ ,  $\dim W = 3$ , and  $L \subset W$ . We use the criterion of Lemma 4.1.4, (iii), applied to the projection  $p: F(1, 3) \rightarrow G(3, N)$  (which is surjective with irreducible fibres) in order to produce a big subset of  $F(1, 3)$  parametrizing smooth intersections. First, the subset  $U' \subset F(1, 3)$  parametrizing such transversal intersections is complement of a countable union of locally closed analytic subsets, since this complement is the image of an analytic set under an analytic map (viz. the projection to  $F(1, 3)$  from the "universal" such intersection, which is a closed analytic subset of  $F(1, 3) \times (X - \{0\})$ ). So by Lemma 4.1.4 (iii), we are reduced to showing that for some big set  $U \subset G(3, N)$ , and any fixed  $W \in U$ , there is a big subset of the fibre  $p^{-1}(W) \cong \mathbb{P}^2$ , such that the points corresponding to smooth intersections form a big subset. There are big subsets  $U_1, U_2$  of  $G(3, N)$  such that (a) for  $W \in U_1$ , the subvariety of  $X - \{0\}$  on which  $W$  fails to generate  $\Omega_{X-\{0\}}^1$  is a smooth surface  $S_W$ , and (b) for  $W \in U_2$ ,  $W$  generates  $\mathcal{O}_{X-\{0\}}$  outside a finite set  $T_W$ . Now take  $U = U_1 \cap U_2$ . For  $W \in U$ ,  $S_W \subset X - \{0\}$  is a smooth complex surface; now apply Lemma 4.1.2 with  $Y = S_W$ ,  $k = 1$ ,  $\mathcal{E} = \mathcal{O}_Y$ ,  $r = 1$  and  $V = W$ ; this yields a big subset of  $\mathbb{P}^2 = p^{-1}(W)$  parametrising smooth hyperplane sections of  $S_W$ , by hyperplanes defined by the vanishing of elements of  $W$ .
7. Let  $(i, j, k)$ ,  $(i', j', k')$  be two sets of indices with exactly two indices in common; then there is a big subset of  $G$  on which  $E_{ijk} \cap E_{i'j'k'}$  is transversal outside a finite set. Let  $F(2, 3) \subset G(2, N) \times G(3, N)$  be the flag variety parametrizing flags of subspaces  $W_1 \subset W_2 \subset \mathbb{C}^N$  with  $\dim W_1 = 2$ ,  $\dim W_2 = 3$ , and let  $p: F(2, 3) \rightarrow G(2, N)$  be the projection. Let

$$F = F(2, 3) \times_{G(2, N)} F(2, 3)$$

be the fibre product, parametrizing pairs of such flags with the same two-dimensional subspace  $W_1$ . The choice of indices  $(i, j, k)$ ,  $(i', j', k')$  gives a morphism  $\psi: G \rightarrow F$  whose image is the complement of the diagonal, hence Zariski open and dense. So it suffices to find a big subset of  $F$  corresponding to transversal intersections.

Let  $U' \subset F$  be the subset of the complement of the diagonal such that for  $(W_1, W_2, W'_2) \in F$  (with  $\dim W_1 = 2$ ,  $\dim W_2 = \dim W'_2 = 3$  and  $W_1 = W_2 \cap W'_2$ ) then the subvariety of  $X - \{0\}$  where neither  $W_2$  nor  $W'_2$  generate  $\Omega_{X-\{0\}}^1$  is a reduced curve (this is the desired transversality condition). Clearly  $U'$  is the complement of a countable union of locally closed analytic subsets. We will use Lemma 4.1.4 (iii) to show that it is big.

Consider the projection  $q: F \rightarrow G(3, N)$  induced by the first projection  $F \rightarrow F(2, 3)$ , followed by the projection  $F(2, 3) \rightarrow G(2, 3)$ , thus  $q(W_1, W_2, W'_2) = W_2$ . For

$W \in G(3, N)$ , the fibre  $q^{-1}(W)$  is an irreducible subvariety of the flag variety  $F(2, 3)$ . There is a big subset of  $G(3, N)$  corresponding to subspaces  $W$  such that the subvariety  $S_W \subset X - \{0\}$ , where  $W$  does not generate  $\Omega_{X-\{0\}}^1$ , is a smooth surface. On  $S_W$  we have that

$$\text{im}(W \otimes_{\mathbb{C}} \mathcal{O}_{S_W} \rightarrow \Omega_{X-\{0\}}^1 \otimes \mathcal{O}_{S_W}) = \mathcal{F}$$

is locally free on  $S_W$  of rank 2.

The fibre of  $F(2, 3) \rightarrow G(3, N)$  over  $W$  is  $\cong \mathbb{P}^2$ , parametrizing two dimensional subspaces of  $W$ ; there is a big subset parametrizing the  $W_1 \subset W$  such that the subvariety of  $S_W$ , where

$$W_1 \otimes_{\mathbb{C}} \mathcal{O}_{S_W} \rightarrow \mathcal{F}$$

has rank  $< 2$  (i.e., is not an isomorphism), is a smooth curve  $T_{W_1}$  (apply Lemma 4.1.2 with  $Y = S_W$ ,  $\mathcal{E} = \mathcal{F}$ , etc). The fibre of  $p_1: F \rightarrow F(2, 3)$  over  $(W_1, W)$  is isomorphic to the projective space of lines in  $\mathbb{C}^N/W_1$ . On  $S_W - T_{W_1}$ , the sheaf

$$\mathcal{E} = \Omega_{X-\{0\}}^1 / \mathcal{F}$$

is an invertible sheaf, generated by space  $\mathbb{C}^N/W_1$  of global sections. Now Lemma 4.1.2 yields a big subset of  $p_1^{-1}(W_1, W)$  such that for  $W_2$  in this subset, the locus where the section

$$(W'_2/W_1) \otimes \mathcal{O}_{S_W} \rightarrow \mathcal{E}$$

vanishes is a smooth curve  $T'$ , meeting  $T_{W_1}$  in a finite set. Finally, since the reduced curve  $T_{W_1}$  is the locus where the sections  $W_1$  do not generate  $\mathcal{F}$ ,

$$W_1 \otimes \mathcal{O}_{S_W} \rightarrow \mathcal{F}$$

has rank 1 along  $T_{W_1}$ , and the cokernel of the above sheaf map is an invertible sheaf on  $T_{W_1}$ . This implies, by local computation, that the locus where the sheaf map

$$W'_2 \otimes \mathcal{O}_{S_W} \rightarrow \Omega_{X-\{0\}}^1 \otimes \mathcal{O}_{S_W}$$

has rank 2 is the union of  $T'$  and the reduced curve  $T_{W_1}$ . This is the desired transversality condition.

8. Let  $(i, j, k)$ ,  $(i', j', k')$  be two sets of indices with exactly one index in common. Then there is a big subset of  $G$  on which  $E_{ijk}$  intersects  $E_{i'j'k'}$  transversally except at a finite set.

This is similar to the previous case 7. Again, one looks at the fibre product  $F$  of the flag variety  $F(1, 3)$  with itself over the projection to  $\mathbb{P}^{N-1}$ . There are maps  $F \rightarrow F(1, 3) \rightarrow G(3, N)$ . By the criterion of Lemma 4.1.4 (iii), it suffices to note that (i) for a big set of  $W$  in  $G(3, N)$ ,  $S_W$  is a smooth surface and (ii) fixing such a  $W$ , for a big subset in the fibre  $\cong \mathbb{P}^2$  of  $F(1, 3) \rightarrow G(3, N)$ , the one dimensional subspace  $W_1$  generates a line sub-bundle of  $\Omega_{X-\{0\}}^1 \otimes \mathcal{O}_{S_W}$ , outside a finite set  $T_{W_1}$  (zero-set of a gen. section of a rank 2 locally free sheaf on  $S_W$ , namely the subsheaf of  $\Omega_{X-\{0\}}^1|_{S_W}$  generated by  $W$ ). (iii) Fix such a flag  $(W_1, W)$ . On the complement of  $T_{W_1}$ , we are left with the problem of finding the zero locus of a "general" two dimensional subspace of  $\mathbb{C}^N/W_1$ , considered as a space of sections generating a locally free sheaf of rank 2 on a surface, which is a smooth curve.

### Acknowledgements

The author would like to thank W C Hsiang. He expresses his gratitude to V Srinivas and A J Parameswaran for the appendix, which is crucial to this paper. He also thanks the referees for valuable comments and criticism.

### References

- [1] Chavel, I, *Eigenvalues in Riemannian Geometry*, (New York: Academic Press) 1984
- [2] Cheeger J, Hodge Theory of Riemannian Pseudomanifolds, *AMS colloq. Publ.* Vol 36 "Geometry of the Laplace Operator"
- [3] Cheeger J, Spectral Geometry of Singular Riemannian Spaces, *J. Diff. Geom.* **18** (1983), 575-657
- [4] Hironaka H, Resolution of singularities of an Algebraic Variety in Characteristic 0, Brandeis Notes, 1962
- [5] Hsiang W-C, Pati V,  $L^2$ -Cohomology of normal algebraic surfaces I. *Inv. Math.* **81** (1985) 395-412
- [6] Nagase M, On the heat operators of normal singular algebraic surfaces. *J. Diff. Geom.* **28** (1988) 37-57
- [7] Pati V,  $L^2$ -Cohomology of algebraic varieties PhD Thesis. Princeton, 1985
- [8] Pati V, The heat trace of singular algebraic threefolds, *J. Diff. Geom.* **37** (1993) 245-261





## The Hoffman–Wielandt inequality in infinite dimensions

RAJENDRA BHATIA and LUDWIG ELSNER\*

Indian Statistical Institute, 7, SJS Sansanwal Marg, New Delhi 110016, India

\* Fakultät für Mathematik, Universität Bielefeld, 4800 Bielefeld, Germany

MS received 1 December 1993

**Abstract.** The Hoffman–Wielandt inequality for the distance between the eigenvalues of two normal matrices is extended to Hilbert–Schmidt operators. Analogues for other norms are obtained in a special case.

**Keywords.** Hoffman–Wielandt inequality; infinite dimensions; Hilbert–Schmidt operators; Schatten  $p$ -norms.

### 1. Introduction

In 1953 Hoffman and Wielandt [13] proved what has now become one of the best-known matrix inequalities. The aim of this paper is to obtain an infinite-dimensional version of this inequality.

Let  $A$  be an  $n \times n$  complex matrix. An  $n$ -tuple  $\{\alpha_1, \dots, \alpha_n\}$  is called an *enumeration* of the eigenvalues of  $A$  if its elements are the eigenvalues of  $A$  each counted as often as its multiplicity. The eigenvalues of  $(A^*A)^{1/2}$  are called the singular values of  $A$  and are denoted as  $s_1(A) \geq s_2(A) \geq \dots \geq s_n(A)$ . We will use the symbol  $\|A\|_2$  to denote what is often called the *Frobenius norm* in the matrix theory literature and the *Hilbert–Schmidt norm* in the operator theory literature. This is defined as

$$\|A\|_2 = (\text{tr } A^*A)^{1/2} = \left[ \sum_{j=1}^n s_j^2(A) \right]^{1/2}. \quad (1)$$

The Hoffman–Wielandt inequality says that if  $A$  and  $B$  are  $n \times n$  normal matrices and if  $\{\alpha_1, \dots, \alpha_n\}$  and  $\{\beta_1, \dots, \beta_n\}$  are enumerations of their eigenvalues, then there exists a permutation  $\pi$  on  $n$  symbols such that

$$\left[ \sum_{i=1}^n |\alpha_i - \beta_{\pi(i)}|^2 \right]^{1/2} \leq \|A - B\|_2. \quad (2)$$

Now let  $\mathcal{H}$  be a complex separable infinite-dimensional Hilbert space. If an operator  $A$  on  $\mathcal{H}$  is compact then the spectrum of  $A$  is a countable set of complex numbers with 0 as the only limit point. All nonzero points in the spectrum are eigenvalues of  $A$  with finite multiplicity. The point 0 may or may not be an eigenvalue of  $A$ , and if it is its multiplicity may be finite or infinite. By an *enumeration* of the eigenvalues of  $A$  we shall mean a sequence  $\{\alpha_1, \alpha_2, \dots\}$  whose terms consist of all the eigenvalues

the nonzero eigenvalues of  $A$  each counted as often as its multiplicity and the term 0 repeated infinitely often.

The singular values of  $A$  are defined as before. Now they are infinite in number. If

$$\|A\|_2 := \left[ \sum_{j=1}^{\infty} s_j^2(A) \right]^{1/2} < \infty \quad (3)$$

the operator  $A$  is said to be in the *Hilbert–Schmidt Class* and the collection of all such operators is denoted as  $\mathcal{H}_2$ .

A bijection  $\pi$  of the set of natural numbers  $\mathbb{N}$  onto itself will be called a *permutation* of  $\mathbb{N}$ .

The following two theorems are infinite-dimensional analogues of the Hoffman–Wielandt Theorem:

**Theorem 1.** *Let  $A$  and  $B$  be normal Hilbert–Schmidt operators and let  $\{\alpha_1, \alpha_2, \dots\}$  and  $\{\beta_1, \beta_2, \dots\}$  be enumerations of their eigenvalues. Then for each  $\varepsilon > 0$  there exists a permutation  $\pi$  of  $\mathbb{N}$  such that*

$$\left[ \sum_{i=1}^{\infty} |\alpha_i - \beta_{\pi(i)}|^2 \right]^{1/2} \leq \|A - B\|_2 + \varepsilon. \quad (4)$$

**Theorem 2.** *Let  $A$  and  $B$  be normal Hilbert–Schmidt operators and let  $\{\alpha'_1, \alpha'_2, \dots\}$  and  $\{\beta'_1, \beta'_2, \dots\}$  be extended enumerations of their eigenvalues. Then there exists a permutation  $\pi$  of  $\mathbb{N}$  such that*

$$\left[ \sum_{i=1}^{\infty} |\alpha'_i - \beta'_{\pi(i)}|^2 \right]^{1/2} \leq \|A - B\|_2. \quad (5)$$

It seems essential to either add an  $\varepsilon$  as in Theorem 1 or to extend the enumerations as in Theorem 2. This point will be discussed in § 2 after the theorems have been proved.

In the special situation when  $A$  and  $B$  are Hermitian our Theorem 2 has already been proved by Markus [16], Friedland [12] and Kato [14], each of whom proved generalisations of this in different directions. Another rather special case was considered by Sakai [18].

The Hilbert–Schmidt norm is one of a family of norms called Schatten  $p$ -norms. These norms are defined as

$$\|A\|_p = \left[ \sum_{j=1}^{\infty} s_j^p(A) \right]^{1/p}, \quad 1 \leq p < \infty \quad (6)$$

$$\|A\|_{\infty} = s_1(A). \quad (7)$$

The class of operators for which  $\|A\|_p$  is finite is an ideal  $\mathcal{H}_p$  in the space of compact operators which itself is denoted as  $\mathcal{H}_{\infty}$ . Basic facts about these norms may be found in several standard texts such as [19].

A problem of much interest in perturbation theory has been that of obtaining analogues of the Hoffman–Wielandt inequality for all these  $p$ -norms (and for the

larger class of symmetric norms). See [3] for a detailed discussion. In both the finite and the infinite dimensional cases this problem has been solved completely when  $A, B$  are both Hermitian (see [2], [14], [15], [16]) and when  $A$  is Hermitian and  $B$  is skew-Hermitian (see [1], [20]). When  $A$  and  $B$  are both unitary this problem has been solved only partially: *sharp* analogues of the inequality (2) are known only for the values  $p = 1$  and  $p = \infty$  and good bounds are known for other values. (See [2], [5], [7], [8], [10]). But when  $A$  and  $B$  are arbitrary normal operators a sharp analogue of (2) for any value of  $p$  other than 2 has not been found *even* in the finite-dimensional case. See [6] and [7] for the known results when  $p = \infty$ .

In this direction we shall prove:

**Theorem 3.** *Let  $A$  be a Hermitian and  $B$  a normal operator, both lying in the Schatten class  $\mathcal{S}_p$  for some  $1 \leq p \leq \infty$ . Let  $\{\alpha'_1, \alpha'_2, \dots\}$  and  $\{\beta'_1, \beta'_2, \dots\}$  be extended enumerations of the eigenvalues of  $A$  and  $B$ . Then there exists a permutation  $\pi$  of  $\mathbb{N}$  such that*

$$\left[ \sum_{i=1}^{\infty} |\alpha'_i - \beta'_{\pi(i)}|^p \right]^{1/p} \leq 2^{2/p-1} \|A - B\|_p \quad \text{if } 1 \leq p \leq 2, \quad (8)$$

and

$$\left[ \sum_{i=1}^{\infty} |\alpha'_i - \beta'_{\pi(i)}|^p \right]^{1/p} \leq 2^{1/2-1/p} \|A - B\|_p \quad \text{if } 2 \leq p \leq \infty, \quad (9)$$

In the finite-dimensional case, the inequality (9) for the special case  $p = \infty$  has been observed earlier. See, e.g., [3, p. 112]. For other  $p$  these results seem to be new even in this case.

## 2. Proofs and remarks

The proofs of Theorems 1 and 2 are both built upon the finite-dimensional case. In the first this involves a straightforward approximation argument, in the second some more intricacies.

*Proof of Theorem 1.* Label the eigenvalues of  $A$  and  $B$  as  $\alpha_1, \alpha_2, \dots$  and  $\beta_1, \beta_2, \dots$  in such a way that

$$|\alpha_1| \geq |\alpha_2| \geq \dots; \quad |\beta_1| \geq |\beta_2| \geq \dots \quad (10)$$

Then choose orthonormal bases  $u_1, u_2, \dots$  and  $v_1, v_2, \dots$  for  $\mathcal{H}$  so that

$$A = \sum_{i=1}^{\infty} \alpha_i u_i u_i^*, \quad B = \sum_{i=1}^{\infty} \beta_i v_i v_i^*. \quad (11)$$

Since  $A$  and  $B$  are both Hilbert-Schmidt operators, given any  $\delta > 0$  we can choose a positive integer  $r$  such that

$$\sum_{i=r+1}^{\infty} |\alpha_i|^2 \leq \delta^2, \quad \sum_{i=r+1}^{\infty} |\beta_i|^2 \leq \delta^2. \quad (12)$$

So, if we define operators  $A_r$  and  $B_r$  as

$$A_r = \sum_{i=1}^r \alpha_i u_i u_i^*, \quad B_r = \sum_{i=1}^r \beta_i v_i v_i^*, \quad (13)$$

then,

$$\|A - A_r\|_2 \leq \delta, \quad \|B - B_r\|_2 \leq \delta. \quad (14)$$

Now consider the linear space spanned by the vectors  $u_1, \dots, u_r$  and  $v_1, \dots, v_r$  together. This is a space of dimension  $s$  where  $r \leq s \leq 2r$ . Call this space  $\mathcal{H}_s$ . The operators  $A_r$  and  $B_r$  both leave  $\mathcal{H}_s$  invariant and vanish on its orthogonal complement. Let  $w_1, \dots, w_s$  be an orthonormal basis for  $\mathcal{H}_s$  in which  $w_j = u_j$  for  $j = 1, 2, \dots, r$ . Then  $A_r w_j = \alpha_j w_j$  for  $1 \leq j \leq r$  and  $A_r w_j = 0$  for  $r+1 \leq j \leq s$ . Define a normal operator  $A_s$  on  $\mathcal{H}_s$  by putting  $A_s w_j = \alpha_j w_j$  for  $1 \leq j \leq s$ . Then note that

$$\|A_s - A_r\|_2^2 = \sum_{j=r+1}^s |\alpha_j|^2 \leq \delta^2. \quad (15)$$

By a similar construction we can define a normal operator  $B_s$  on  $\mathcal{H}_s$  which has eigenvalues  $\beta_1, \dots, \beta_s$  and is such that

$$\|B_s - B_r\|_2 \leq \delta. \quad (16)$$

Now apply the Hoffman–Wielandt Theorem to the operators  $A_s$  and  $B_s$  on the finite-dimensional space  $\mathcal{H}_s$ . This gives a permutation  $\pi$  of the set  $\{1, 2, \dots, s\}$  such that

$$\sum_{j=1}^s |\alpha_j - \beta_{\pi(j)}|^2 \leq \|A_s - B_s\|_2^2. \quad (17)$$

Now extend this permutation  $\pi$  to all of  $\mathbb{N}$  by defining  $\pi(j) = j$  if  $j > s$ . Then the inequalities (12), (14), (15), (16) and (17) together give

$$\sum_{j=1}^{\infty} |\alpha_j - \beta_{\pi(j)}|^2 \leq (\|A - B\|_2 + 4\delta)^2 + 4\delta^2.$$

Since  $\delta$  was arbitrary this proves the theorem. ■

*Proof of Theorem 2.* Once again label the eigenvalues of  $A$  and  $B$  as in (10). Define extended enumerations  $\{\alpha'_i\}$ ,  $\{\beta'_i\}$  of eigenvalues of  $A$  and  $B$  as the two sequences whose terms are

$$\begin{aligned} \alpha'_{2i-1} &= \alpha_i, & \alpha'_{2i} &= 0, & i &= 1, 2, \dots, \\ \beta'_{2i-1} &= \beta_i, & \beta'_{2i} &= 0, & i &= 1, 2, \dots \end{aligned} \quad (18)$$

By a slight modification of the argument used in proving Theorem 1 we can find a sequence  $\varepsilon_n$  of positive numbers and a sequence  $\pi_n$  of permutations of  $\mathbb{N}$  such that

$$\sum_{i=1}^{\infty} |\alpha'_i - \beta'_{\pi_n(i)}|^2 \leq \|A - B\|_2^2 + \varepsilon_n^2, \quad (19)$$

and

$$\lim \varepsilon_n = 0. \quad (20)$$

To see this adopt the same notations as in the proof of Theorem 1 up to the inequality (14). Now let  $\mathcal{H}_n$  be any subspace of dimension  $n = 2r$  which contains all the vectors  $u_1, \dots, u_r, v_1, \dots, v_r$ . The operators  $A_r$  and  $B_r$  both leave  $\mathcal{H}_n$  invariant and their restrictions to this space have eigenvalues  $\alpha'_i, \beta'_i, i = 1, 2, \dots, n$ . So, by the Hoffman-Wielandt Theorem there exists a permutation  $\pi_n$  of  $\{1, 2, \dots, n\}$  such that

$$\sum_{i=1}^n |\alpha'_i - \beta'_{\pi_n(i)}|^2 \leq \|A_r - B_r\|^2 \leq (\|A - B\| + 2\delta)^2.$$

Extend the permutation  $\pi_n$  to all of  $\mathbb{N}$  by putting  $\pi_n(j) = j$  if  $j > n$  and define  $\varepsilon_n$  via  $\delta$  to get (19) and (20). Let

$$v_n = \pi_n^{-1}, \quad n = 1, 2, \dots \quad (21)$$

We now construct a permutation  $\pi$  of  $\mathbb{N}$  that will satisfy (5). To do this we will describe a procedure that defines  $\pi$  and its inverse  $v$  by successively assigning values to  $\pi(1), v(1), \pi(2), v(2), \dots$ . At the same time a subsequence of the sequence  $\{\pi_n\}$  of permutations defined in the preceding paragraph is chosen. The procedure is described below in the form of an algorithm. This has two steps  $\alpha$  and  $\beta$  to be run alternately and in each of these three mutually exclusive choices have to be made.

For  $i = 1, 2, \dots$ , do

$\alpha$  Look successively at the following three options, do as instructed, then go to  $\beta$ :

- (I) (void if  $i = 1$ ). If  $i = v(j)$  for some  $j < i$  define  $\pi(i) = j$ .
- (II) If the set  $\{\pi_n(i) : n = 1, 2, \dots\}$  is bounded let  $j$  be the minimal number which occurs infinitely often in this set. Define  $\pi(i) = j$ . Replace  $\{\pi_n\}$  by a subsequence, denoted again by  $\{\pi_n\}$ , such that now  $\pi_n(i) = j$  for all  $n$ .
- (III) If the set  $\{\pi_n(i) : n = 1, 2, \dots\}$  is unbounded let  $j$  be the smallest even number which has not yet been called  $\pi(k)$  for any  $k < i$ . Define  $\pi(i) = j$ . (Note that in this case  $\lim_{n \rightarrow \infty} \beta'_{\pi_n(i)} = 0$  and we have defined  $\pi$  in such a way that  $\beta'_{\pi(i)} = 0$ .)

$\beta$  Look successively at the following three options, do as instructed, then go back to  $\alpha$  with  $i + 1$  in place of  $i$ :

- (IV) If  $i = \pi(j)$  for some  $j \leq i$  define  $v(i) = j$ .
- (V) If the set  $\{v_n(i) : n = 1, 2, \dots\}$  is bounded let  $j$  be the minimal number which occurs infinitely often in this set. Define  $v(i) = j$ . Replace  $\{v_n\}$  by a subsequence, denoted again by  $\{v_n\}$ , such that now  $v_n(i) = j$  for all  $n$ . This also gives a new subsequence of  $\{\pi_n\}$  if we put  $\pi_n = v_n^{-1}$ .
- (IV) If the set  $\{v_n(i) : n = 1, 2, \dots\}$  is unbounded let  $j$  be the smallest even number that has not yet been called  $v(k)$  for any  $k < i$ . Define  $v(i) = j$ . (Note in this case we had  $\lim_{n \rightarrow \infty} \alpha'_{v_n(i)} = 0$  and we have defined  $v$  in such a way that  $\alpha'_{v(i)} = 0$ .)

We claim that the permutation  $\pi$  defined above satisfies the inequality (5). For this it is enough to show that for every positive integer  $N$  we have

$$\sum_{i=1}^N |\alpha'_i - \beta'_{\pi(i)}|^2 \leq \|A - B\|_2^2. \quad (22)$$

Let  $\Phi_N = \{\pi_1, \pi_2, \dots\}$  be the subsequence of the original sequence  $\{\pi_n\}$  obtained after running  $N$  steps of  $\alpha$  and  $\beta$  in the above procedure. We will split the set  $\{1, 2, \dots, N\}$  into three disjoint subsets  $S_1, S_2$  and  $S_3$  by separating indices according to what happened to them in the above algorithm. These sets are defined as

$$S_1 = \{i: 1 \leq i \leq N, \exists \pi_m \in \Phi_N \text{ such that } \pi(i) = \pi_m(i)\}.$$

Note that if  $i \in S_1$  then by (II) and (V) in the above construction  $\pi(i) = \pi_m(i)$  for all  $\pi_m \in \Phi_N$ .

$$S_2 = \{i: 1 \leq i \leq N, \pi(i) \text{ was defined by (III) above}\}.$$

Note that

$$\beta'_{\pi(i)} = \lim_{n \rightarrow \infty} \beta'_{\pi_n(i)} = 0 \quad \text{if } i \in S_2. \quad (23)$$

$S_3 = \{i: 1 \leq i \leq N, i \text{ was defined as } i = v(j) \text{ for some } j \leq i \text{ by (VI) above}\}$ . Note that

$$\alpha'_i = \lim_{n \rightarrow \infty} \alpha'_{v_n(\pi(i))} = 0 \quad \text{if } i \in S_3. \quad (24)$$

Now for any element  $\pi_n$  of  $\Phi_N$  we can use the above splitting to write

$$\begin{aligned} \sum_{i=1}^N |\alpha'_i - \beta'_{\pi(i)}|^2 &= \sum_{i \in S_1} |\alpha'_i - \beta'_{\pi_n(i)}|^2 + \sum_{i \in S_2} |\alpha'_i|^2 + \sum_{i \in S_3} |\beta'_{\pi(i)}|^2 \\ &= \left[ \sum_{i \in S_1} |\alpha'_i - \beta'_{\pi_n(i)}|^2 + \sum_{i \in S_2} |\alpha'_i - \beta'_{\pi_n(i)}|^2 + \sum_{i \in S_3} |\alpha'_{v_n(\pi(i))} - \beta'_{\pi(i)}|^2 \right] \\ &\quad + \sum_{i \in S_2} \{|\alpha'_i|^2 - |\alpha'_i - \beta'_{\pi_n(i)}|^2\} + \sum_{i \in S_3} \{|\beta'_{\pi(i)}|^2 - |\alpha'_{v_n(\pi(i))} - \beta'_{\pi(i)}|^2\}. \end{aligned} \quad (25)$$

As  $n \rightarrow \infty$  the last two sums in (25) go to zero, since both are finite sums of terms going to zero. The limit of the three sums inside the square brackets can be written as

$$\lim_{n \rightarrow \infty} \sum_{i=1}^N |\alpha'_i - \beta'_{\pi_n(i)}|^2.$$

This is bounded above by

$$\lim_{n \rightarrow \infty} \sum_{i=1}^{\infty} |\alpha'_i - \beta'_{\pi_n(i)}|^2.$$

Hence, the inequality (22) follows from (19) and (20). ■

The difference between the finite-dimensional and the infinite-dimensional case arises because of the fact that the *unitary orbit* of an operator  $A$  defined as the set  $\{UAU^*: U \text{ unitary}\}$  is closed in the former case but not always in the latter. The following simple example illustrating this phenomenon was provided to us by Peter Rosenthal.

*Example.* Let  $A$  be the normal operator given by

$$A = \text{diag}\left(1, \frac{1}{2}, \frac{1}{3}, \dots\right).$$

For  $n = 1, 2, \dots$ , let

$$A_n = \text{diag}\left(\frac{1}{n}, 1, \frac{1}{2}, \dots, \frac{1}{n-1}, \frac{1}{n+1}, \frac{1}{n+2}, \dots\right)$$

Then each  $A_n$  is in the unitary orbit of  $A$ . However,  $A_n$  converges (in the Hilbert-Schmidt norm topology) to  $B$  where

$$B = \text{diag}\left(0, 1, \frac{1}{2}, \frac{1}{3}, \dots\right)$$

and  $B$  is not in the unitary orbit of  $A$ . By the same argument we can find a sequence of operators in the unitary orbit of  $A$  which converges to a diagonal operator having arbitrarily many zeroes on the diagonal.

One way to interpret the inequality (2) is that it gives a lower bound for the distance between the unitary orbits of two diagonal matrices. In the infinite-dimensional case such orbits are not closed. So, we should seek a lower bound for the distance between their *closures*. Such a bound is provided by Theorem 2.

The other, more standard, interpretation of (2) is that it gives an upper bound for the distance between the eigenvalues of two normal matrices. This distance is a metric on the space of unordered  $n$ -tuples of complex numbers. More precisely, consider the space  $\mathbb{C}^n$  with the Euclidean norm  $\|\cdot\|_2$ . Let  $\Pi_n$  be the group of permutations on  $n$  indices. For  $x \in \mathbb{C}^n$  let  $x(\pi)$  be the vector whose coordinates are obtained by applying the permutation  $\pi$  to the coordinates of  $x$ . Calling two such vectors equivalent let  $\tilde{x}$  be the equivalence class of  $x$ . Let  $\tilde{\mathbb{C}}^n = \mathbb{C}^n / \Pi_n$  be the space of such equivalence classes. Then this is a metric space with the metric

$$d(\tilde{x}, \tilde{y}) = \min_{\pi} \|x - y(\pi)\|_2.$$

Since the eigenvalues of an  $n \times n$  matrix are known only up to a permutation it is natural to identify them with a point in the space  $\tilde{\mathbb{C}}^n$ . The inequality (2) then gives a bound for the distance between the eigenvalues of two normal matrices  $A$  and  $B$  in terms of the distance between  $A$  and  $B$ . Now when  $A$  is a Hilbert-Schmidt operator we have to replace the space  $\mathbb{C}^n$  in the above discussion by the space  $l_2$ . Let  $\Pi$  denote the set of all bijections of the set of natural numbers onto itself.

Consider the space  $\tilde{l}_2 = l_2 / \Pi$ . The eigenvalues of  $A$  can be identified with a point in this space. We can now define for  $\tilde{x}, \tilde{y}$  in this space

$$d(\tilde{x}, \tilde{y}) = \inf_{\pi} \|x - y(\pi)\|_2.$$

However, the example given above also shows that this does not give a metric on  $\tilde{l}_2$ . It only gives a *pseudometric*. Indeed, given any  $x$  in  $l_2$  we can find a  $y$  which has

which  $d(\tilde{x}, \tilde{y}) = 0$ . The quotient space  $\tilde{l}_2/d$  with respect to this pseudometric is a metric space. To identify this space let  $l'_2$  be the subset of  $l_2$  consisting of vectors with infinitely many zero entries. For each  $x \in l_2$  let  $x' = (x_1, 0, x_2, 0, \dots)$ . Then  $x' \in l'_2$ . Let  $\tilde{x}'$  be the image of this point in  $\tilde{l}_2 = l'_2/\Pi$ . Now define

$$d(\tilde{x}, \tilde{y}') = \inf_{\pi} \|x' - y'(\pi)\|_2.$$

It can be seen that this defines a metric on the space  $\tilde{l}_2$ . It would be most natural to use this metric to measure the distance between the eigenvalues of two Hilbert-Schmidt operators. Theorem 2 is then seen to be the natural extension of the finite-dimensional Hoffman-Weilandt Theorem.

Since 0 is *always* an accumulation point of the eigenvalues of a compact operator, in any case there is good reason to include it with infinite multiplicity in a count of the eigenvalues.

Now we recall briefly some of the known results for the special class of Hermitian operators. Let  $A$  and  $B$  be  $n \times n$  Hermitian matrices with eigenvalues enumerated as  $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_n$  and  $\beta_1 \geq \beta_2 \geq \dots \geq \beta_n$  respectively. Then we have

$$\left[ \sum_{i=1}^n |\alpha_i - \beta_i|^p \right]^{1/p} \leq \|A - B\|_p \quad \text{for } 1 \leq p \leq \infty. \quad (26)$$

This is a consequence of a theorem of Lidskii and Wielandt. See [3, Chapter 3]. This theorem was extended to infinite dimensions by Markus [16]. If  $A$  is a compact Hermitian operator associate with it a doubly infinite sequence  $\{\alpha_{\pm j}; j \in \mathbb{N}\}$  satisfying the following conditions

- (i)  $\alpha_1 \geq \alpha_2 \geq \dots \geq 0$ ,  
 $\alpha_{-1} \leq \alpha_{-2} \leq \dots \leq 0$ ;
- (ii) if  $A$  has infinitely many positive and infinitely many negative eigenvalues then the sequence  $\{\alpha_{\pm j}\}$  contains only these numbers each repeated as often as its multiplicity as an eigenvalues of  $A$  (0 is not included in the sequence in this case even if it is an eigenvalue of  $A$ );
- (iii) if  $A$  has only finitely many positive eigenvalues then the sequence  $\{\alpha_j\}$  contains these repeated according to their multiplicities and an infinite number of zero terms; and if  $A$  has only finitely many negative eigenvalues then the sequence  $\{\alpha_{-j}\}$  contains these repeated according to their multiplicities and an infinite number of zero terms.

With this notation Markus proves a result from which it follows that if  $A$  and  $B$  are compact Hermitian operators and if  $\{\alpha_{\pm j}\}$  and  $\{\beta_{\pm j}\}$  are sequences associated with them according to the above rules then

$$\left[ \sum_{j=1}^{\infty} \{|\alpha_j - \beta_j|^p + |\alpha_{-j} - \beta_{-j}|^p\} \right]^{1/p} \leq \|A - B\|_p \quad \text{for } 1 \leq p \leq \infty. \quad (27)$$

This device of adding zeroes to make both the positive and the negative eigenvalues of  $A$  infinite in number achieves exactly what our extended enumeration did. One can easily see that the "optimal matching" of the eigenvalues of  $A$  and  $B$  is achieved



by the pairing in (27). If both  $A$  and  $B$  have infinitely many positive and negative eigenvalues then extending the enumerations by adding zeroes does not affect the sums involved. So, for the Hermitian case our Theorem 2 is included in this result of Markus. The  $p = 2$  case of (27) is also proved in Friedland [12].

Kato [14] has proved a similar result in the more general situation when  $A$  and  $B$  are any two bounded Hermitian operators whose difference is compact. Let  $\sigma(A)$  denote the spectrum of a Hermitian operator  $A$ . An isolated point of  $\sigma(A)$  is always an eigenvalue of  $A$ ; if it has finite multiplicity call it a *discrete eigenvalue*. Let  $\sigma_d(A)$  be the collection of all such points. The complement of  $\sigma_d(A)$  in  $\sigma(A)$  is called the *essential spectrum* of  $A$  and is denoted as  $\sigma_{\text{ess}}(A)$ . Eigenvalues of  $A$  that have infinite multiplicity are in  $\sigma_{\text{ess}}(A)$  whether they are isolated points of  $\sigma(A)$  or not. The set  $\sigma_{\text{ess}}(A)$  is a closed subset of  $\mathbb{R}$  and so its complement in  $\mathbb{R}$  is a countable union of open intervals  $I_n$ . Kato defines an *extended enumeration of discrete eigenvalues* of  $A$  to be a sequence  $\{\alpha_j\}$  with the following properties

- (i) every discrete eigenvalue of  $A$  appears in this sequence as often as its multiplicity,
- (ii) all other points of the sequence  $\{\alpha_j\}$ , belong to the set of boundary points of the open intervals  $I_n$  mentioned above.

We should add here an explanatory note. An extended enumeration  $\{\alpha_j\}$  according to the above definition need not include all the boundary points of all the intervals  $I_n$  and those that are included may be counted as often as one wishes.

With this definition Kato proves that if  $A$  and  $B$  are Hermitian operators such that  $A - B$  is compact then there exist extended enumerations  $\{\alpha_j\}$  and  $\{\beta_j\}$  of discrete eigenvalues of  $A$  and  $B$  such that

$$\left[ \sum_{j=1}^{\infty} |\alpha_j - \beta_j|^p \right]^{1/p} \leq \|A - B\|_p \text{ for } 1 \leq p \leq \infty. \quad (28)$$

The result of Markus can be derived from this.

We should add that all the inequalities (26)–(28) are true for the larger class of symmetric norms.

Our Theorem 3 is proved using the above results for the Hermitian case. We will need the following facts. Let

$$T = T_1 + iT_2 = \frac{T + T^*}{2} + i \frac{T - T^*}{2} \quad (29)$$

be the Cartesian decomposition of any operator  $T$ . Then.

$$\|T\|_2^2 = \|T_1\|_2^2 + \|T_2\|_2^2. \quad (30)$$

$$\|T_1\|_{\infty} \leq \|T\|_{\infty}, \|T_2\|_{\infty} \leq \|T\|_{\infty}. \quad (31)$$

If  $T$  is normal then the eigenvalues of  $T_1$  and  $T_2$  are the real and the imaginary parts of the eigenvalues of  $T$ . We will use the Clarkson–McCarthy inequalities which say that if  $T$  and  $S$  are in the Schatten class  $\mathcal{S}_p$  then

$$2(\|T\|_p^p + \|S\|_p^p) \leq \|T + S\|_p^p + \|T - S\|_p^p \quad \text{for } 2 \leq p \leq \infty \quad (32)$$

$$2^{p-1}(\|T\|_p^p + \|S\|_p^p) \leq \|T + S\|_p^p + \|T - S\|_p^p \quad \text{for } 1 \leq p \leq 2. \quad (33)$$

See [9] or [19]. We will also use the elementary inequalities:

$$|x + iy|^p \leq 2^{p/2-1}(|x|^p + |y|^p) \quad \text{for } 2 \leq p \leq \infty, \quad (34)$$

$$|x + iy|^p \leq |x|^p + |y|^p \quad \text{for } 1 \leq p \leq 2, \quad (35)$$

valid for all real numbers of  $x$  and  $y$ .

*Proof of Theorem 3.* Let  $B = B_1 + iB_2$  be the Cartesian decomposition of  $B$ . We shall apply the inequality (27) to the Hermitian operators  $A$  and  $B_1$ . Let us represent extended enumerations of eigenvalues of  $A$  and  $B$  in the form of doubly infinite sequences  $\{\alpha_{\pm j}\}$  and  $\{\beta_{\pm j}\}$  in which

$$\alpha_1 \geq \alpha_2 \geq \dots \geq 0, \quad \alpha_{-1} \leq \alpha_{-2} \leq \dots \leq 0;$$

$$\operatorname{Re} \beta_1 \geq \operatorname{Re} \beta_2 \geq \dots \geq 0, \quad \operatorname{Re} \beta_{-1} \leq \operatorname{Re} \beta_{-2} \leq \dots \leq 0.$$

In all summations the index  $j$  will run over positive and negative integers.

The case  $p = 2$  is specially simple. We have from (30) and (27)

$$\begin{aligned} \|A - B\|_2^2 &= \|A - B_1\|_2^2 + \|B_2\|_2^2 \\ &\geq \sum_j |\alpha_j - \operatorname{Re} \beta_j|^2 + \sum_j |\operatorname{Im} \beta_j|^2 \\ &= \sum_j |\alpha_j - \beta_j|^2, \end{aligned}$$

which is the desired inequality.

The case  $p = \infty$  is equally simple. Use (31) instead of (30). For each  $j$  we have

$$\begin{aligned} |\alpha_j - \beta_j|^2 &= |\alpha_j - \operatorname{Re} \beta_j|^2 + |\operatorname{Im} \beta_j|^2 \\ &\leq \|A - B_1\|_\infty^2 + \|B_2\|_\infty^2 \\ &\leq 2\|A - (B_1 + iB_2)\|_\infty^2 \\ &= 2\|A - B\|_\infty^2. \end{aligned}$$

For  $2 \leq p < \infty$  use (32) and (34) together with (27) to get

$$\begin{aligned} 2^{1-p/2} \sum_j |\alpha_j - \beta_j|^p &\leq \sum_j |\alpha_j - \operatorname{Re} \beta_j|^p + \sum_j |\operatorname{Im} \beta_j|^p \\ &\leq \|A - B_1\|_p^p + \|B_2\|_p^p \\ &\leq \frac{1}{2} \{ \|A - B_1 + iB_2\|_p^p + \|A - B_1 - iB_2\|_p^p \} \\ &= \frac{1}{2} \{ \|A - B^*\|_p^p + \|A - B\|_p^p \} \\ &= \|A - B\|_p^p, \end{aligned}$$

which is the desired inequality.

For  $1 \leq p < 2$  use (33) and (35) together with (27) to get the result

Sakai [18] has proved a rather special case of the above Theorem. He proves it for  $p = 2$  assuming that  $A$  and  $B_1$  are both positive operators. In the special case when  $A$  is Hermitian and  $B$  skew-Hermitian stronger inequalities for all  $p$ -norms have been obtained by Ando and Bhatia [1].

We end with some remarks about results which can be easily proved using the same ideas.

Hoffman and Wielandt also proved an inequality complementary to (2). There exists a permutation  $\pi$  such that

$$\|A - B\|_2 \leq \left[ \sum_{i=1}^n |\alpha_i - \beta_{\pi(i)}|^2 \right]^{1/2}.$$

Such complementary inequalities for (4) and (5) can also be obtained.

Let  $(A^{(1)}, \dots, A^{(m)})$  be an  $m$ -tuple of pairwise commuting compact normal operators in  $\mathcal{H}$ . Then there exists an orthonormal basis  $e_j$ ,  $j = 1, 2, \dots$ , such that each  $e_j$  is a simultaneous eigenvector for all  $A^{(k)}$ ,  $1 \leq k \leq m$ . Let  $A^{(k)}e_j = \lambda_j^{(k)}e_j$ ,  $1 \leq k \leq m$ . The points  $(\lambda_j^{(1)}, \dots, \lambda_j^{(m)})$  in the space  $\mathbb{C}^m$ ,  $j = 1, 2, \dots$ , can be called the *joint eigenvalues* of the tuple  $(A^{(1)}, \dots, A^{(m)})$ . The set of these points together with the point 0 in  $\mathbb{C}^m$  coincides with the *Taylor spectrum* and the *Harte spectrum* in this case. See, e.g., [17]. In [4] and [11] it was shown that the Hoffman-Wielandt inequality (2) can be extended to give or bound for the distance between the joint eigenvalues of two commuting  $m$ -tuples of normal matrices. Following the same ideas our Theorems 1 and 2 can also be generalised to commuting  $m$ -tuples of normal Hilbert-Schmidt operators.

A version of Theorem 3 when  $A$  and  $B$  are not compact but  $A - B$  is, can be proved using Kato's Theorem and the ideas of our proof. Note that  $A - B = A - B_1 - iB_2$ . So both  $A - B_1$  and  $B_2$  are compact if  $A - B$  is. An extended enumeration of discrete eigenvalues of  $B$  should now mean a sequence  $\{\beta_i\}$  such that  $\{\operatorname{Re} \beta_j\}$  is such an enumeration for  $B_1$  in the sense of Kato.

In [9] the Clarkson-McCarthy inequalities are generalised to all unitarily invariant norms. These can be used to obtain some results extending Theorem 3 to such norms.

## Acknowledgements

The first author thanks the University of Bielefeld and the second author thanks the Indian Statistical Institute for visits during which this work was done. Both authors thank Sonderforschungsbereich 343 and DAE India, for financial support.

## References

- [1] Ando T and Bhatia R, Eigenvalue inequalities associated with the Cartesian decomposition, *Linear Multilinear Algebra*, **22** (1987) 133-147
- [2] Bhatia R, Analysis of spectral variation and some inequalities, *Trans. Am. Math. Soc.*, **272** (1982) 323-331
- [3] Bhatia R, *Perturbation Bounds for Matrix Eigenvalues* (Essex Longman) (1987)
- [4] Bhatia R and Bhattacharyya T, A generalisation of the Hoffman-Wielandt theorem, *Linear Algebra Appl.*, **179** (1993) 11-17

- [5] Bhatia R and Davis C, A bound for the spectral variation of a unitary operator, *Linear Multilinear Algebra*, **15** (1984) 71–76
- [6] Bhatia R, Davis C and Koosis P, An extremal problem in Fourier analysis with applications to operator theory, *J. Funct. Anal.*, **82** (1989) 138–150
- [7] Bhatia R, Davis C and McIntosh A, Perturbation of spectral subspaces and solution of linear operator equations, *Linear Algebra Appl.* **52** (1983) 45–67
- [8] Bhatia R and Holbrook J A R, Short normal paths and spectral variation, *Proc. Am. Math. Soc.*, **94** (1985) 377–382
- [9] Bhatia R and Holbrook J A R, On the Clarkson–McCarthy inequalities, *Math. Ann.* **281** (1988) 7–12
- [10] Bhatia R and Sinha K B, A unitary analogue of Kato's theorem on variation of discrete spectra, *Lett. Math. Phys.*, **15** (1988) 201–204
- [11] Elsner L, A note on the Hoffman–Wielandt theorem, *Linear Algebra Appl.*, **182** (1993) 235–237
- [12] Friedland S, Inverse eigenvalue problems, *Linear Algebra Appl.*, **17** (1977) 15–51
- [13] Hoffman A J and Wielandt H W, The variation of the spectrum of a normal matrix, *Duke Math. J.*, **20** (1953) 37–39
- [14] Kato T, Variation of discrete spectra, *Commun. Math. Phys.*, **111** (1987) 501–504
- [15] Lidskii V B, The proper values of the sum and product of symmetric matrices, *Dokl. Akad. Nauk, SSSR*, **75** (1950) 769–772
- [16] Markus A S, The eigen and singular values of the sum and product of linear operators, *Russ. Math. Surv.*, **19** (1964) 92–120
- [17] McIntosh A, Pryde A and Ricker W, Comparison of joint spectra for certain classes of commuting operators, *Stud. Math.*, **88** (1988) 23–36
- [18] Sakai Y, Continuous versions of an inequality due to Hoffman and Wielandt, *Linear Algebra Appl.*, **71** (1985) 283–287
- [19] Simon B, *Trace Ideals and Their Applications*, (Cambridge: University Press), (1979)
- [20] Sunder V S, Distance between normal operators, *Proc. Am. Math. Soc.*, **84** (1982) 483–484

## Rigidity problem for lattices in solvable Lie groups

A N STARKOV

Department of Mathematics, Moscow State University, 117234 Moscow, Russia

MS received 4 June 1993; revised 7 October 1993

**Abstract.** The paper concerns rigidity problem for lattices in simply connected solvable Lie groups. A lattice  $\Gamma \subset G$  is said to be rigid if for any isomorphism  $\phi: \Gamma \rightarrow \Gamma'$  with another lattice  $\Gamma' \subset G$  there exists an automorphism  $\hat{\phi}: G \rightarrow G$  which extends  $\phi$ . An effective rigidity criterion is proved which generalizes well-known rigidity theorems due to Malcev and Saito. New examples of rigid and nonrigid lattices are constructed. In particular, we construct: a) rigid lattice  $\Gamma \subset G$  which is not Zariski dense in the adjoint representation of  $G$ , b) Zariski dense lattice  $\Gamma \subset G$  which is not rigid, c) rigid virtually nilpotent lattice  $\Gamma$  in a solvable nonnilpotent Lie group  $G$ .

**Keywords.** Rigidity problem; Zariski density; lattice; Lie groups.

### Introduction

Recall the basic definition. A lattice  $\Gamma$  in a Lie group  $G$  is said to be *rigid* (weakly rigid) in  $G$  if for any isomorphism  $f: \Gamma \rightarrow \Gamma'$  of  $\Gamma$  with another lattice  $\Gamma' \subset G$  (respectively for any automorphism  $f: \Gamma \rightarrow \Gamma$ ) there exists an automorphism  $\hat{f}: G \rightarrow G$  such that  $\hat{f}|_{\Gamma} = f$ .

The famous Mostow–Margulis–Prasad theorem (see, for instance, [8]) gives a sufficient condition for lattices in a broad class of semisimple Lie groups to be rigid. The solvable case seems to be less well studied. To formulate known results in this case let us give a classification of solvable groups mostly following [8].

The class of simply connected solvable Lie groups is divided into subclasses  $(N)$ ,  $(I)$ ,  $(R)$ ,  $(E)$ ,  $(A)$  in such a way that  $(N) \subset (I)$ ,  $(N) \subset (R) \subset (E) \subset (A)$  and  $(I) \cap (A) = (N)$ . Here  $(N)$  is the class of nilpotent groups,  $(R)$  and  $(I)$  are so-called classes of groups of “real” and “imaginary” types and  $(E)$  is the class of “exponential” groups. These classes are well known. We introduce the new class  $(A)$  named after L Auslander, whose paper [1] presents the first nontrivial example of Lie group of such a type with a lattice (see § 1 for exact definitions). This class is the maximal opposite class to the class  $(I)$ :

**Theorem 5.1.** *Let  $\Gamma$  be a lattice in a connected simply connected solvable Lie group  $G$ . Then there exist normal connected subgroups  $G_I$  and  $G_A$  of types  $(I)$  and  $(A)$  respectively such that*

$$(1) \quad G = G_I G_A,$$

$$(2) \quad \Gamma = (\Gamma \cap G_I)(\Gamma \cap G_A), \text{ and}$$

$$(3) \quad G \cap G_A = N, \text{ where } N \text{ is the nilradical of } G.$$

Now let us recall known results on the rigidity problem in the solvable case. The first nontrivial result in this field is Malcev's theorem [15] on the rigidity of all lattices in nilpotent Lie groups. Saito [21] generalized this result to all lattices in solvable groups of  $(R)$ -type.

Consider the semidirect product  $G = \mathbf{R} \cdot \mathbf{R}^2$ , where the line  $\mathbf{R}$  acts on the plane  $\mathbf{R}^2$  by orthogonal rotations with the kernel  $2\pi\mathbf{Z} \subset \mathbf{R}$ . Then  $G$  is of  $(I)$ -type and the abelian lattice  $\Gamma = 2\pi\mathbf{Z} \times \mathbf{Z}^2$  is not even weakly rigid in  $G$  (it suffices to permute  $2\pi\mathbf{Z} \subset \mathbf{R}$  and a cyclic subgroup  $\mathbf{Z}^1 \subset \mathbf{Z}^2 \subset \mathbf{R}^2$ ). However, it can be proved (see example 2.8) that the lattice  $\Gamma' = (\frac{1}{2}\pi\mathbf{Z}) \cdot \mathbf{Z}^2$  is weakly rigid (but not rigid) in  $G$ .

Milovanov [16] has constructed the first example of a nonrigid lattice in a Lie group of  $(E)$ -type. Thus, Saito rigidity theorem cannot be generalized to all lattices in solvable groups of  $(I)$  or  $(E)$ -type. Note that important results on the rigidity of Zariski-dense arithmetic lattice in algebraic groups were proved by Platonov and Milovanov [19].

We can give an effective criterion for a lattice  $\Gamma$  in a simple connected solvable Lie group  $G$  to be rigid. We use the important injection  $G \subset \text{Hol}(U) = \text{Aut}(U) \cdot U$  of  $G$  into the real algebraic group  $\text{Hol}(U)$ , where  $U$  is the so-called unipotent hull of  $G$ . The concept of the unipotent hull  $U$  was introduced by Hochschild and Mostow [10] and in an essentially different way by Auslander and Green [4]. However, the idea of the injection of a simply connected solvable group into an algebraic group goes back to Malcev [14]. The injection  $G \subset \text{Hol}(U)$  was intensively used in the theory of solvmanifolds (see, for instance, [11], [6], [2]) and plays an important role in the rigidity problem.

According to Milovanov [17], any simply connected solvable Lie group  $G'$  with a lattice  $\Gamma' \simeq \Gamma \subset G$  may be injected into  $\text{Hol}(U)$  in such a way that  $\Gamma \subset G'$ . Let  $S(G, \Gamma)$  be the family of all groups  $G' \subset \text{Hol}(U)$  such that  $\Gamma \subset G'$  and  $G' \simeq G$ . This family is obviously divided into subfamilies  $C(G', \Gamma)$  of all groups  $G'_m$  of the form  $G'_m = mG'm^{-1}$ , where  $m \in \text{Hol}(U)$  commutes with  $\Gamma$ . We prove the following rigidity criterion:

**Theorem 3.6.** *Let  $\Gamma$  be a lattice in a connected simply connected solvable Lie group  $G$ . Then the following conditions are equivalent:*

- (1)  $\Gamma$  is rigid in  $G$ .
- (2)  $S(G, \Gamma) = C(G, \Gamma)$ .

With the help of this criterion we prove in §4 that the example of a lattice  $\Gamma$  in a Lie group  $G$  of  $(A)$ -type constructed in [1] is rigid. This shows that rigid lattices exist outside the class  $(R)$  of solvable groups. It may be noted that this rigid lattice  $\Gamma \subset G$  is Zariski-dense in  $G$  (relative to the algebraic structure of  $\text{Hol}(U)$ ). In fact, all lattices in solvable groups of  $(R)$ -type are also Zariski-dense (and rigid). However, we construct in example 6.1 a rigid lattice which is not Zariski-dense and in example 6.2 a nonrigid Zariski-dense lattice. These examples are heavily based on the construction from [1] as well.

Also, it may be noted that the Lie groups  $G$  in these examples are splittable (i.e.  $G = A \cdot N$ , where an abelian subgroup  $A \subset G$  acts on the nilradical  $N \subset G$  by semisimple automorphisms). We prove (proposition 5.2) that a splittable solvable group with a rigid lattice should be of  $(A)$ -type. In particular, there are no rigid lattices in splittable

nonnilpotent groups of  $(I)$ -type. However, a nonsplittable group of  $(I)$ -type may have a rigid lattice. This example is constructed in § 8 with the help of conjecture of Grunewald and Segal [9] on the structure of the algebraic group  $\text{Aut}(n)$  for a nilpotent Lie algebra  $\eta$ , proved by Bryant and Groves [7].

At last, we consider in § 7 the closely related problem of rigidity of  $\Gamma \subset G$  under deformations. We prove that the weakly rigid lattice  $\Gamma' = (\frac{1}{2}\pi\mathbf{Z}) \cdot \mathbf{Z}^2$  in the simplest Lie group  $G = \mathbf{R} \cdot \mathbf{R}^2$  of  $(I)$ -type is rigid under deformations but not rigid. Also, we prove that every Zariski-dense lattice  $\Gamma \subset G$  is rigid under deformations. In particular, the Zariski-dense lattice from example 6.2 is rigid under deformations but not weakly rigid.

## 1. Classification of solvable Lie groups and semisimple splitting construction

Let  $G$  be a connected simply connected solvable Lie group. The type of the Lie group  $G$  is determined by the set of eigenvalues of the adjoint operators  $\text{Ad}(g)$ ,  $g \in G$  on the Lie algebra  $L(G)$ .

### DEFINITION 1.1

$G$  is said to be of  $(I)$ -type (from "imaginary") if every eigenvalue  $\lambda$  of any operator  $\text{Ad}(g)$ ,  $g \in G$  lies on the unit circle:  $|\lambda| = 1$ .

### DEFINITION 1.2

$G$  is said to be of  $(R)$ -type (from "real") if every eigenvalue  $\lambda$  of any operator  $\text{Ad}(g)$ ,  $g \in G$  is purely real.

### DEFINITION 1.3

$G$  is said to be of  $(E)$ -type (from "exponential") if every eigenvalue  $\lambda \neq 1$  of any operator  $\text{Ad}(g)$ ,  $g \in G$  lies out of the unit circle:  $|\lambda| \neq 1$ .

### DEFINITION 1.4

$G$  is said to be of  $(A)$ -type if any operator  $\text{Ad}(g)$ ,  $g \in G$  is either unipotent or has at least one eigenvalue  $\lambda$  out of the unit circle:  $|\lambda| \neq 1$ .

The first three kinds of groups are well known. It should be noted, however, that sometimes solvable groups of  $(I)$ -type are called as  $(R)$ -type (from "rotation", see [4]) and the type  $(R)$  in our classification had no common name before. Our classification corresponds to that of [8] and seems to be more natural. The last type of a solvable Lie groups is new. It is named after L Auslander whose paper [1] presents the first nontrivial example of a Lie group  $G$  of such a type with a lattice  $\Gamma \subset G$ .

Let  $(N)$  denote the class of nilpotent Lie groups. Then we have obvious relationships between the classes of a solvable Lie groups:  $(N) \subset (R) \subset (E) \subset (A)$ ,  $(N) \subset (I)$  and  $(I) \cap (A) = (N)$ . We see later (Theorem 5.1) that the class  $(A)$  is the maximal opposite class to the class  $(I)$ .

It is well-known that every connected subgroup and quotient group of a Lie group  $G$  of  $(N)$ ,  $(R)$ ,  $(E)$  or  $(I)$ -type has the same type. The class  $(E)$  is completely characterized by the property that a group of  $(E)$ -type has no nonnilpotent subgroups of  $(I)$ -type

(i.e., has no subgroups of  $(I)\backslash(N)$ -type). Also,  $G$  is of  $(E)$ -type if and only if the exponential map  $\exp: L(G) \rightarrow G$  is a homeomorphism (see, for instance, [5]). The class  $(A)$  may be characterized by the property that there are no normal subgroups of  $(I)\backslash(N)$ -type of a solvable Lie group of  $(A)$ -type. However, a Lie group of  $(A)$ -type may have a subgroup or quotient group of  $(I)$ -type (see example 4.1) and, consequently, the class  $(A)$  is not closed under such natural operations as localisation to a subgroup or taking a quotient.

The classification of solvable Lie groups may be given also with the help of the construction  $G_S$  of the semisimple splitting (Malcev splitting) of a solvable Lie group  $G$ . This construction enables us to define a very important injection  $G \subset \text{Hol}(U)$  of a solvable Lie group  $G$  into the algebraic group  $\text{Hol}(U) = \text{Aut}(U) \cdot U$ , where  $U$  is the nilradical in  $G_S$ , called the unipotent hull of  $G$  ([10, 11]) or the nilshadow of  $G$  ([4]). It should be noted that the concept of the unipotent hull  $U$  of a solvable Lie group  $G$  was introduced by Hochschild and Mostow [10], while the injection  $G \subset \text{Hol}(U)$  was for the first time constructed and used in solvable group theory by Auslander *et al* [3, 6]. However, the idea of the injection of a solvable Lie group into some algebraic group goes back to Malcev [14].

The fastest way to introduce the splitting  $G_S$  for a solvable group  $G$  is the non-constructive definition due to [4].

#### DEFINITION 1.5

Let  $T_G$  be a connected abelian subgroup of semisimple elements of the algebraic group  $\text{Aut}(G)$  of automorphisms of a simply connected solvable Lie group  $G$  and let  $U$  be the nilradical in the group  $G_S = T_G \cdot G$ . If

- (1)  $G_S = T_G \cdot U$  and
- (2)  $G_S = GU$

then  $U$  is called the *unipotent hull* of  $G$  and  $G_S$  is called the *semisimple splitting* of  $G$ .

Denote by  $p: G_S = T_G \cdot U \rightarrow T_G$  the projection of  $G_S$  onto  $T_G$  and by  $\pi: G_S = T_G \cdot U \rightarrow U$  the projection of  $G_S$  onto  $U$ . It follows from the definition that the restriction  $p: G \rightarrow T_G$  is an epimorphism of groups and the restriction  $\pi: G \rightarrow U$  is a diffeomorphism of manifolds. Also, the nilradical  $N$  of the group  $G$  has the form  $N = (G \cap U)_0$ .

Clearly the group  $T_G$  consists of nontrivial semisimple automorphisms of  $U$  and we may view  $T_G$  as a subgroup of the algebraic group  $\text{Aut}(U)$ . This allows us to consider the injection  $G \subset G_S = T_G \cdot U \subset \text{Hol}(U) = \text{Aut}(U) \cdot U$ . The group  $U$  being a connected simply connected nilpotent Lie group admits the structure of a real algebraic group and so does the holomorph  $\text{Hol}(U) = \text{Aut}(U) \cdot (U)$  ([2]). The injection  $G \subset \text{Hol}(U)$  allows us to use methods of the algebraic group theory and will play an important role in our paper.

The structure of the semisimple splitting  $G_S$  may be successfully demonstrated on the example of so-called splittable solvable Lie groups.



To construct the group  $G_S$  for a splittable group  $G = A \cdot N$  let us denote by  $*$ :  $A \rightarrow \text{Aut}(N)$  the natural homomorphism. Then the image  $*(A) = A^*$  is an abelian subgroup of  $\text{Aut}(N)$  consisting of semisimple elements and we may view  $A^*$  as a subgroup in  $\text{Aut}(G)$ . Let  $T_G = A^*$  and  $G_S = T_G \cdot G$ . Then the nilradical  $U$  of  $G_S$  has the form  $U = N \times \Delta$ , where  $\Delta$  is the "antidiagonal" in  $A^* \times A$ , i.e. the set of all elements of the form  $(a^*, a^{-1})$ ,  $a \in A$ . Clearly  $A^* \times A = A^* \times \Delta = A\Delta$  and the group  $G_S = A^* \cdot G = A^* \cdot U$  is the semisimple splitting of  $G$ . We have proved the following.

### PROPOSITION 1.7

*Let  $G = A \cdot N$  be a splittable solvable Lie group. Then the semisimple splitting  $G_S$  of  $G$  has the form  $G_S = T_G \cdot U$ , where  $T_G = A^*$ ,  $U = N \times \Delta$  and  $\Delta$  is the "antidiagonal" in  $A^* \times A$ .*

In principle, the subgroup  $T_G \subset \text{Aut}(U)$  is not closed even in the Euclidean topology. However, the following strengthened version of Mostow structural theorem [12] holds.

**Theorem 1.8 [2].** *If a solvable Lie group  $G$  has a lattice  $\Gamma \subset G$  then*

- (1)  $T_G$  is a closed subgroup of  $\text{Aut}(U)$  in the Euclidean topology.
- (2) The image  $p(\Gamma) \subset T_G$  of  $\Gamma$  under the projection  $p: G_S = T_G \cdot U \rightarrow T_G$  is a lattice in  $T_G$ .

Let  $A(H)$  denote the algebraic hull of a subgroup  $H \subset \text{Hol}(U)$ ; by that we mean the smallest algebraic subgroup of  $\text{Hol}(U)$  containing  $H$ . Then  $G$  is Zariski-dense in  $G_S$ , i.e.  $A(G) = A(G_S) = A(T_G) \cdot U$ . The following theorem is a consequence of Mostow's fundamental results [11] as well.

**Theorem 1.9 [2].** *Let  $\Gamma$  be a lattice in a simply connected solvable Lie group  $G$ . Then  $U \subset A(\Gamma)$  and the quotient space  $A(G)/A(\Gamma)$  is compact.*

It can be shown [3] that the semisimple splitting  $G_S$  and the unipotent hull  $U$  are uniquely determined by  $G$ . However, there is always some arbitrariness in the choice of  $T_G$  (all such subgroups are conjugate by elements of  $U$ ). The group  $G_S$  is by definition splittable and it is known [3] that every element  $g \in G$  admits a Jordan decomposition, i.e. for each  $g \in G$  there is a choice of  $T_G$  such that  $g = p(g)\pi(g)$ , where the semisimple element  $p(g) \in T_G$  and the unipotent element  $\pi(g) \in U$  commute. Since all subgroups  $T_G$  in  $G_S$  are conjugate, the eigenvalues of the adjoint operator  $\text{Ad}$  for the elements  $p(g) \in T_G$  and  $g \in G$  coincide independently of the choice of  $T_G$ .

This remark enables us to give a simple classification of solvable Lie groups in terms of the semisimple splitting. Clearly, a Lie group  $G$  is nilpotent if and only if  $T_G = 1$ . For the rest of the paper we assume that a solvable Lie group  $G$  has a lattice  $\Gamma \subset G$ , and hence the subgroup  $T_G \subset \text{Aut}(U)$  is closed. Put  $N = (G \cap U)_0$  and  $U_I = A(G \cap U)$ . Then  $N$  is the nilradical in  $G$  and we have  $N \subset U_I \subset U$ .

### PROPOSITION 1.10

*The following conditions are equivalent:*

*Proof.*  $1 \Leftrightarrow 2$ . Since  $T_G$  is a closed subgroup of  $\text{Aut}(U)$  of semisimple elements, it follows that  $T_G$  is compact if and only if all eigenvalues of every operator  $\text{Ad}(a), a \in T_G$  lie on the unit circle.

$2 \Leftrightarrow 3$ . Since  $T_G = p(G) = G/(G \cap U)$  and  $G$  is diffeomorphic to  $U$ ,  $T_G$  is compact if and only if  $U/(G \cap U)$  is compact, i.e.  $G \cap U$  is a uniform subgroup of the simply connected nilpotent group  $U$ , and hence  $U = A(G \cap U) = U_I$ .

In order to give the criterion for a solvable Lie group  $G$  to be of (R) or (A)-type let us recall that every algebraic abelian subgroup  $A$  of semisimple elements has the decomposition  $A = T \times D$ , where  $T$  is the maximal compact subgroup of  $A$  and  $D$  is a simply connected algebraic subgroup such that every element  $a \in D$  has only purely real eigenvalues. This decomposition corresponds to the decomposition of a  $\mathbb{R}$ -algebraic abelian group over  $\mathbb{C}$  into its isotropic and anisotropic parts.

### PROPOSITION 1.11

*$G$  is of (R)-type if and only if the algebraic abelian group  $A(T_G)$  is simply connected.*

*Proof.* If  $G$  is of (R)-type, then all eigenvalues of every operator  $\text{Ad}(a), a \in T_G$  are purely real and, consequently, so are the eigenvalues of every operator  $\text{Ad}(a), a \in A(T_G)$ . Hence,  $A(T_G)$  is simply connected.

On the other hand, if  $A(T_G)$  is simply connected, it follows that all eigenvalues of every operator  $\text{Ad}(a), a \in A(T_G)$  are purely real and so are all eigenvalues of every operator  $\text{Ad}(g), g \in G$ , i.e.  $G$  is of (R)-type.

### PROPOSITION 1.12

*The following conditions are equivalent.*

- (1)  $G$  is of (A)-type.
- (2)  $T_G$  is a simply connected subgroup of  $\text{Aut}(U)$ .
- (3)  $N = U_I$ .

*Proof.*  $1 \Leftrightarrow 2$ . Clearly the closed subgroup  $T_G \subset \text{Aut}(U)$  has no compact subgroups if and only if every semisimple operator  $\text{Ad}(a), a \in T_G, a \neq 1$  has at least one eigenvalue out of the unit circle, i.e.  $G$  is of (A)-type.

$2 \Leftrightarrow 3$ . Since  $T_G = p(G) = G/(G \cap U)$ , it follows that  $T_G$  is simply connected if and only if  $G \cap U$  is connected, i.e.  $N = (G \cap U)_0 = A(G \cap U) = U_I$ .

## 2. The rigidity of lattices in Lie groups

Let us begin our discussion of the rigidity problem by giving necessary definitions.

### DEFINITION 2.1

A lattice  $\Gamma \subset G$  is said to be *rigid* if for any isomorphism  $\phi: \Gamma \rightarrow \Gamma'$  of  $\Gamma$  onto another lattice  $\Gamma' \subset G$  there exists an automorphism  $\hat{\phi}: G \rightarrow G$  such that  $\hat{\phi}|_{\Gamma} = \phi$ .

## DEFINITION 2.2

A lattice  $\Gamma \subset G$  is said to be *weakly rigid* if for any automorphism  $\phi: \Gamma \rightarrow \Gamma$  there exists an automorphism  $\hat{\phi}: G \rightarrow G$  such that  $\hat{\phi}|_{\Gamma} \equiv \phi$ .

Clearly the rigidity property for a lattice is stronger than weak rigidity one. The example 2.8 provides us with a weakly rigid but not rigid lattice  $\Gamma \subset G$ . To begin with, let us recall some well-known results on the rigidity problem for lattices in solvable Lie groups.

**Theorem 2.3** (Malcev, [15]). *All lattices in a nilpotent Lie group are rigid.*

**Theorem 2.4** (Saito, [21]). *All lattices in a solvable Lie group of (R)-type are rigid.*

To construct some elementary examples of nonrigid lattices let us prove the necessary condition for a lattice in a splittable Lie group to be rigid.

## PROPOSITION 2.5

Let  $G = A \cdot V$  be a simply connected splittable Lie group, where  $V$  is a vector space and  $A$  is an abelian group of a semisimple automorphisms of  $V$ . Let  $\Gamma$  be a lattice in  $G$  and  $\phi: \Gamma \rightarrow \Gamma'$  be an isomorphism of  $\Gamma$  onto another lattice  $\Gamma' \subset G$ . Then  $\phi$  may be extended to an automorphism  $\hat{\phi}: G \rightarrow G$  only if

- (1)  $\phi(\Gamma \cap V) \subset V$ , and hence  $\phi|_{\Gamma \cap V}$  may be extended to an automorphism  $\phi^*: V \rightarrow V$ , and
- (2) The element  $\phi^* \in \text{Aut}(V)$  normalizes the image  $A^* \subset \text{Aut}(V)$  of the abelian group  $A \subset G$ .

*Proof.* Since  $V$  is the nilradical of  $G$ , it follows that  $\hat{\phi}(V) = V$  and, in particular,  $\phi(\Gamma \cap V) \subset \hat{\phi}(V) = V$  and the necessity of the first condition is obvious.

Let  $\phi^* \in \text{Aut}(V)$  be the extension of the isomorphism  $\phi: \Gamma \cap V \rightarrow \Gamma' \cap V$ . Then  $\hat{\phi}|_V \equiv \phi^*$ . Each element  $a \in A$  induces the automorphism  $a^* \in \text{Aut}(V)$  given as  $a^*(v) = av a^{-1}$  for every  $v \in V$ . We have  $\hat{\phi}(a)^*(v) = \hat{\phi}(a)v\hat{\phi}^{-1}(a) = \hat{\phi}(a\phi^{*-1}(v)a^{-1}) = \phi^*(a\phi^{*-1}(v)a^{-1}) = (\phi^*a^*\phi^{*-1})(v)$ . Hence,  $\hat{\phi}(a)^* = \phi^*a^*\phi^{*-1}$  for each  $a \in A$ . Since all maximal abelian subgroups of semisimple elements of a splittable group are conjugate, it follows that  $\hat{\phi}(A) = v_0 A v_0^{-1}$  for some  $v_0 \in V$ . But since  $V$  is abelian, we have  $\hat{\phi}(A)^* = A^*$  and therefore  $A^* = \hat{\phi}(A)^* = \phi^* A^* \phi^{*-1}$ .

Now let us construct the first example of a nonrigid lattice.

**Example 2.6.** Let  $\mathbf{R}$  denote the real line. Put  $G = S(\mathbf{R}) \cdot \mathbf{R}^2$ , where the one-parameter group  $S(\mathbf{R})$  acts on the plane  $\mathbf{R}^2$  by orthogonal rotations with the kernel  $S(2\pi\mathbf{Z})$ . Then  $G$  is a simply connected splittable Lie group of (I)-type. Let the lattice  $\Gamma \subset G$  have the form  $\Gamma = S(2\pi\mathbf{Z}) \times \mathbf{Z}^2$ , where  $\mathbf{Z}^2$  is an integer lattice in  $\mathbf{R}^2$ . Since every element  $\phi^* \in SL(2, \mathbf{Z})$  induces an automorphism of  $\mathbf{Z}^2$  and commutes with the trivial action of  $S(2\pi\mathbf{Z})$ , it determines an automorphism  $\phi$  of the lattice  $\Gamma$ . The image of the group  $S(\mathbf{R})$  in  $\text{Aut}(\mathbf{R}^2) = GL(2, \mathbf{R})$  is nothing but the circle  $SO(2) = S(\mathbf{R})/S(2\pi\mathbf{Z})$ . If the lattice  $\Gamma$  were weakly rigid then by the second condition of the proposition 2.5 the group  $SO(2)$  would be normalized by every element  $\phi^*$  of  $SL(2, \mathbf{Z})$ . This contradiction proves that  $\Gamma$  is not weakly rigid in  $G$ .

It is not hard to construct an automorphism  $\phi: \Gamma \rightarrow \Gamma$  such that the first condition of the proposition 2.5 is not satisfied as well. For instance, it suffices to permute the group  $S(2\pi\mathbf{Z})$  and some cyclic subgroup  $\mathbf{Z}' \subset \mathbf{Z}^2$ .

*Example 2.7.* Even if the conditions of the proposition 2.5 are satisfied, an isomorphism  $\phi: \Gamma \rightarrow \Gamma'$  may not be extended to an automorphism  $\hat{\phi}: G \rightarrow G$ . In fact, let  $G$  be as above. Then the subgroup  $S(2\pi\mathbf{Z}) \times \mathbf{R}^2$  is abelian and we may put the lattice  $\Gamma' \subset G$  under the form  $\Gamma' = Z(a) \times \mathbf{Z}^2$ , where  $Z(a)$  is the cyclic group generated by the element  $a = s(2\pi) \times v$ , with  $v \in \mathbf{R}^2$ ,  $v \neq 1$ . Then  $\Gamma' \cap S(2\pi\mathbf{Z}) = 1$  but there exists an isomorphism  $\phi: \Gamma \rightarrow \Gamma'$  of the lattice  $\Gamma = S(2\pi\mathbf{Z}) \times \mathbf{Z}^2$  onto the lattice  $\Gamma' = Z(a) \times \mathbf{Z}^2$  such that  $\phi|_{\mathbf{Z}^2} \equiv Id$ . Note that both conditions of the proposition 2.5 are satisfied, since  $\phi$  induces the trivial automorphism  $\phi^* = 1 \in \text{Aut}(\mathbf{R}^2)$ . If there were an extension  $\hat{\phi}: G \rightarrow G$  then the image  $\hat{\phi}(S(2\pi\mathbf{Z}))$  of the centre  $S(2\pi\mathbf{Z}) \subset G$  would be the unipotent subgroup  $Z(a) \subset G$ .

*Example 2.8.* Now let us construct an example of a weakly rigid and nonrigid lattice  $\Gamma$  in the Lie group  $G$  as above. Put  $\Gamma = S(\frac{1}{2}\pi\mathbf{Z}) \cdot \mathbf{Z}^2$  and let  $\phi: \Gamma \rightarrow G$  be the homomorphism defined by the rule  $\phi|_{\mathbf{Z}^2} \equiv Id$  and  $\phi(s(\pi/2)) = s(5\pi/2)$ . Since the actions of the elements  $s(\pi/2)$  and  $s(5\pi/2)$  on  $\mathbf{R}^2$  coincide, it follows that  $\phi$  defines an isomorphism  $\phi: \Gamma \rightarrow \Gamma'$  of  $\Gamma$  onto another lattice  $\Gamma' \subset \Gamma \subset G$ . Note that  $\phi$  is not an automorphism of  $\Gamma$ , since  $\Gamma'$  is a proper subgroup of  $\Gamma$ . If there were an extension  $\hat{\phi}: G \rightarrow G$ , then the restriction of  $\hat{\phi}$  to the center  $S(2\pi\mathbf{Z})$  of the group  $G$  would define an automorphism of  $S(2\pi\mathbf{Z})$ . But the restriction of  $\phi$  to  $S(2\pi\mathbf{Z})$  is the multiplication by 5. This proves nonrigidity of  $\Gamma$ .

However, the lattice  $\Gamma$  is weakly rigid in  $G$ . To prove this let us consider an arbitrary automorphism  $\phi: \Gamma \rightarrow \Gamma$ . Note that the restriction of  $\phi$  to the centre  $S(2\pi\mathbf{Z})$  of  $\Gamma$  defines an automorphism of  $S(2\pi\mathbf{Z})$ . Then there are two possibilities:

1)  $\phi(s(2\pi)) = s(2\pi)$ . In this case  $\phi(s(\pi/2)) = s(\pi/2)v$ , with some  $v \in \mathbf{R}^2$ . It is easy to verify that for every  $v \in \mathbf{R}^2$  there exists an element  $m \in \mathbf{R}^2$  such that  $s(\pi/2)v = ms(\pi/2)m^{-1}$ . On the other hand, the restriction of  $\phi$  on the commutant  $\mathbf{Z}^2$  of  $\Gamma$  defines an element  $\phi^* \in \text{Aut}(\mathbf{Z}^2)$ . Since the actions of  $s(\pi/2)$  and  $ms(\pi/2)m^{-1}$  on  $\mathbf{R}^2$  are represented by the matrix

$$a^* = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

it follows that  $a^*$  commutes with  $\phi^* \in \text{Aut}(\mathbf{Z}^2)$ . But the centralizer of  $a^*$  in  $\text{Aut}(\mathbf{Z}^2)$  coincides with the group  $SO(2, \mathbf{Z}) = S(\frac{1}{2}\pi\mathbf{Z})^*$ , and hence  $\phi^*(n) = bnb^{-1}$  for every  $n \in \mathbf{R}^2$  and some  $b \in S(\frac{1}{2}\pi\mathbf{Z})$ . Put  $\hat{\phi}(an) = mam^{-1}\phi^*(n)$  for all  $a \in S(\mathbf{R})$ ,  $n \in \mathbf{R}^2$ . Then  $\hat{\phi}$  is an automorphism of  $G$  such that  $\hat{\phi}|_{\Gamma} \equiv \phi$ .

2)  $\phi(s(2\pi)) = s(-2\pi)$ . Then  $\phi(s(\pi/2)) = ms(-\pi/2)m^{-1}$  for some  $m \in \mathbf{R}^2$  and the element  $\phi^* \in \text{Aut}(\mathbf{Z}^2)$  should satisfy the condition  $\phi^*a^*\phi^{*-1} = a^{*-1}$ , because the action of  $\phi(s(\pi/2))$  on  $\mathbf{R}^2$  coincides with the action of  $a^{*-1}$ . Therefore,

$$a^* = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

Thus, we have constructed an example of a weakly rigid and nonrigid lattice in a solvable Lie group of  $(I)$ -type. Now let us recall the known example of a nonrigid lattice in a solvable Lie group of  $(E)$ -type.

**Example 2.9** (Milovanov, [16]). Let  $\alpha \in \mathbf{R}$  be such that  $e^\alpha + e^{-\alpha} = 3$ . Consider the action of a one-paramater group  $D(\mathbf{R})$  on a vector space  $\mathbf{R}^4 = \mathbf{R}_u^2 \times \mathbf{R}_v^2$  defined by the rule  $D(t)(u_1, u_2, v_1, v_2) = (e^{\alpha t} u_1, e^{\alpha t} u_2, e^{-\alpha t} v_1, e^{-\alpha t} v_2)$ . Then there exists a suitable choice of a basis in  $\mathbf{R}^4$  such that the action  $D(1)$  is represented by the integer matrix

$$\begin{pmatrix} 2 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

having the eigenvalues  $\lambda_{1,3} = e^\alpha, \lambda_{2,4} = e^{-\alpha}$ . Thus, we may form the discrete group  $\Gamma = D(\mathbf{Z}) \cdot \mathbf{Z}^4$ , where  $\mathbf{Z}^4$  is an integer lattice in  $\mathbf{R}^4$  relative to the suitable basis in  $\mathbf{R}^4$  and  $D(\mathbf{Z})$  is the cyclic group generated by the element  $D(1)$ . Put  $G = D(\mathbf{R}) \cdot \mathbf{R}^4$ . Then  $G$  is a simply connected solvable Lie group of  $(R)$ -type and by theorem 2.4,  $\Gamma$  is a rigid lattice in  $G$ .

Now let us consider the action of the circle  $SO(2)$  on  $\mathbf{R}_u^2$  by orthogonal rotations together with the trivial action on  $\mathbf{R}_v^2$ . This defines the action  $\{O(t), t \in \mathbf{R}\}$  of  $SO(2)$  on  $\mathbf{R}^4$ , which commutes with the action  $D(\mathbf{R})$ . In an explicit form  $O(t)(u_1, u_2, v_1, v_2) = (\cos(t)u_1 + \sin(t)u_2, -\sin(t)u_1 + \cos(t)u_2, v_1, v_2)$ . Let  $\{A(t) = D(t) \times O(2\pi t), t \in \mathbf{R}\}$  be the diagonal subgroup in the cylinder  $D(\mathbf{R}) \times O(\mathbf{R})$ . Put  $G' = A(\mathbf{R}) \cdot \mathbf{R}^4$ . Then  $G'$  is a "distortion" of  $G$  and contains the same lattice  $\Gamma$ . Clearly  $G'$  is a simply connected solvable Lie group of  $(E)$ -type but not of  $(R)$ -type. Let us prove that  $\Gamma$  is nonrigid in  $G'$ .

In fact, every automorphism of the plane  $\mathbf{R}_u^2$  commutes with the action of  $D(\mathbf{R})$  on  $\mathbf{R}^4$  and determines an automorphism  $\hat{\phi}: G \rightarrow G$  such that  $\hat{\phi}(\Gamma) = \Gamma' \subset D(\mathbf{Z}) \cdot \mathbf{R}^4 \subset G'$ . But by proposition 2.5, the restriction  $\hat{\phi}: \Gamma \rightarrow \Gamma'$  may have no extension to  $G'$ , since the automorphism of  $\mathbf{R}_u^2$  may not normalize the circle  $SO(2) \subset \text{Aut}(\mathbf{R}_u^2)$ .

Note that the lattice  $\Gamma$  is Zariski-dense and rigid in  $G$ , while it is not Zariski-dense and rigid in  $G'$ . Thus, the rigidity theorem 2.4 cannot be generalized from the class  $(R)$  of solvable Lie groups to the class  $(E)$ .

### 3. The rigidity criterion

To state our rigidity criterion let us first recall the following fundamental result due to Mostow (see [11] or [20, Theorem 4.41]) in a form convenient to us.

**Theorem 3.1** [17]. *Let  $H$  and  $H'$  be uniform subgroups in simply connected solvable Lie groups  $G$  and  $G'$  respectively. Then for every isomorphism  $\phi: H \rightarrow H'$  there exists the unique algebraic extension  $\hat{\phi}: A(H) \rightarrow A(H')$ , where  $A(H) \subset A(G) \subset \text{Hol}(U)$ ,  $A(H') \subset A(G') \subset \text{Hol}(U')$  and  $U, U'$  are the unipotent hulls of  $G, G'$  respectively.*

Note that there may be many nonalgebraic extensions  $\hat{\phi}: A(H) \rightarrow A(H')$  but they are not interesting to us. The following two particular cases of this theorem will be of great importance to us: when  $H = G$  and when  $H = \Gamma$  is a lattice in  $G$ .

Using theorem 3.1, Milovanov [17] proved that any simply connected solvable

In particular, the unipotent hull  $U$  is uniquely determined by  $\Gamma$ . As a consequence, Milovanov proved that up to an isomorphism there are at most countably many simply connected solvable groups  $G'$  with a lattice isomorphic to  $\Gamma$ .

Henceforth, everything take place inside the big algebraic group  $\text{Hol}(U) = \text{Aut}(U) \cdot U$ . Let us begin our discussion of the rigidity problem by trying to extend an isomorphism  $\phi: \Gamma \rightarrow \Gamma'$  of lattices  $\Gamma, \Gamma' \subset G$  to an automorphism of  $G$ . By theorem 3.1, any isomorphism  $\phi: \Gamma \rightarrow \Gamma'$  uniquely determines an algebraic isomorphism  $\tilde{\phi}: A(\Gamma) \rightarrow A(\Gamma')$ . Let us study the structure of  $\tilde{\phi}$ . Denote by  $\phi_* \in \text{Aut}(U)$  the restriction  $\tilde{\phi}|_U$ .

### DEFINITION 3.2

Let  $\phi_* \in \text{Aut}(U)$ . Then the isomorphism  $\tilde{\phi}: \text{Hol}(U) \rightarrow \text{Hol}(U)$  given as  $\tilde{\phi}(an) = (\phi_* a \phi_*^{-1}) \phi_*(n)$ , for all  $a \in \text{Aut}(U)$ ,  $n \in U$ , is said to be *canonical*.

Clearly, an algebraic isomorphism  $\tilde{\phi}: A(\Gamma) \rightarrow A(\Gamma')$  may not coincide with the restriction  $\tilde{\phi}|_{A(\Gamma)}$ . However,  $\tilde{\phi}$  is always conjugate to  $\tilde{\phi}|_{A(\Gamma)}$ . To prove this let us denote by  $u_* \in \text{Int}(U) \subset \text{Aut}(U)$  the inner automorphism of  $U$  generated by any element  $u \in U$ .

### PROPOSITION 3.3

Let  $\tilde{\phi}: A(\Gamma) \rightarrow A(\Gamma')$  be an algebraic isomorphism and let  $A(\Gamma) = A(T_\Gamma) \cdot U$ , where  $T_\Gamma = p(\Gamma) \subset A(T_\Gamma) \subset \text{Aut}(U)$ . Then there exists an element  $u \in U$  such that  $\tilde{\phi}(an) = (uu_*^{-1} \phi_*) a (\phi_*^{-1} u_* u^{-1}) \phi_*(n)$  for all  $a \in A(T_\Gamma)$ ,  $n \in U$ .

*Proof.* For all  $a \in A(T_\Gamma)$ ,  $n \in U$ , we have  $\tilde{\phi}(a) \tilde{\phi}(n) \tilde{\phi}(a)^{-1} = \tilde{\phi}(ana^{-1}) = \phi_*(ana^{-1}) = (\phi_* a) (n) = (\phi_* a \phi_*^{-1}) (\phi_*(n)) = (\phi_* a \phi_*^{-1}) \phi_*(n) (\phi_* a^{-1} \phi_*^{-1})$ .

On the other hand, let us consider the decomposition  $\text{Aut}(U) = P \cdot V$  of the algebraic group  $\text{Aut}(U)$  into some maximal reductive subgroup  $P \supset A(T_\Gamma)$  and the unipotent radical  $V$ . Then  $\text{Hol}(U) = P \cdot (V \cdot U)$  is the decomposition of  $\text{Hol}(U)$  into the maximal reductive subgroup  $P$  and the unipotent radical  $V \cdot U$ . Since all maximal reductive subgroups of the algebraic group  $\text{Hol}(U)$  are conjugate by elements of  $V \cdot U$ , it follows that there exist a subgroup  $C \subset P$  and elements  $v \in V, u \in U$  such that  $\tilde{\phi}(A(T_\Gamma)) = uvCv^{-1}u^{-1}$ . Hence, for any  $a \in A(T_\Gamma)$  there exists  $c \in C$  such that for all  $n \in U$  we have  $\tilde{\phi}(a) \tilde{\phi}(n) \tilde{\phi}(a)^{-1} = (uvcv^{-1}u^{-1}) \phi_*(n) (uvcv^{-1}v^{-1}u^{-1}) = (u_* vcv^{-1}u_*^{-1}) \phi_*(n) (u_* vc^{-1}v^{-1}u_*^{-1})$ .

Since the elements  $\phi_* a \phi_*^{-1}$  and  $u_* vcv^{-1}u_*^{-1}$  lie in  $\text{Aut}(U)$  and their actions on  $U$  coincide, we conclude that  $\phi_* a \phi_*^{-1} = u_* vcv^{-1}u_*^{-1}$  and  $\tilde{\phi}(a) = uvcv^{-1}u^{-1} = (uu_*^{-1} \phi_*) a (\phi_*^{-1} u_* u^{-1})$ . This completes the proof of the proposition.

### COROLLARY 3.4

Let  $\psi: \text{Hol}(U) \rightarrow \text{Hol}(U)$  denote the canonical isomorphism determined by the element  $u_*^{-1} \phi_* \in \text{Aut}(U)$ . Then  $\tilde{\phi}(g) = u\psi(g)u^{-1}$ , for all  $g \in A(\Gamma)$ .

Hence, we may define at least one algebraic extension  $\phi^*: \text{Hol}(U) \rightarrow \text{Hol}(U)$  given as  $\phi^*(g) = u\psi(g)u^{-1}$ , for all  $g \in \text{Hol}(U)$ . Clearly, this extension is not unique. Indeed, let  $z \in \text{Hol}(U)$  commute with  $\Gamma$ . Then  $z$  determines another algebraic extension  $\phi_z^*: \text{Hol}(U) \rightarrow \text{Hol}(U)$  given as  $\phi_z^*(g) = z\phi^*(g)z^{-1}$ , for all  $g \in \text{Hol}(U)$ . It is easy to prove that any such element  $z \in \text{Hol}(U)$  has the form  $z = mm_*^{-1}$ , where  $m \in Z_U(T_\Gamma)$ , and  $Z_U(T_\Gamma) \subset U$  denotes the centralizer of the reductive group  $T_\Gamma$  in  $U$ . Fortunately, this

describes all possible algebraic extensions  $\phi^*: \text{Hol}(U) \rightarrow \text{Hol}(U)$  of the given isomorphism  $\phi: \Gamma \rightarrow \Gamma'$ :

### COROLLARY 3.5

Let  $\phi^*: \text{Hol}(U) \rightarrow \text{Hol}(U)$  be an algebraic isomorphism such that  $\phi^*|_{A(\Gamma)} \equiv \text{Id}$  and let  $A(\Gamma) = A(T_\Gamma) \cdot U$ . Then there exists an element  $m \in Z_U(T_\Gamma)$  such that  $\phi^*(an) = (mm_*^{-1})a(m_*m_*^{-1})n$  for all  $a \in \text{Aut}(U)$ ,  $n \in U$ .

*Proof.* Since  $\phi^*|_{A(\Gamma)} \equiv \text{Id}$ , it follows that  $\phi^*|_U \equiv \text{Id}$  and  $mm_*^{-1}$  commutes with the reductive group  $T_\Gamma$ . But  $m \in U$ , while  $m_* \in \text{Aut}(U)$ , and hence  $m \in Z_U(T_\Gamma)$ .

Now we are able to give the criterion for a lattice  $\Gamma \subset G$  to be rigid. Let  $S(G, \Gamma)$  denote the family of all solvable Lie groups  $G' \subset \text{Hol}(U)$  such that  $\Gamma \subset G'$  and  $G' \simeq G$ . It follows from the previous proposition that this family is divided into continuous subfamilies  $C(G', \Gamma)$  of all groups  $G'_z$  of the form  $G'_z = zG'z^{-1}$ , where  $z \in \text{Hol}(U)$  commutes with  $\Gamma$ .

**Theorem 3.6** (The rigidity criterion). *Let  $\Gamma$  be a lattice in a connected simply connected solvable Lie group  $G$ . Then the following conditions are equivalent:*

- (1)  $\Gamma$  is rigid in  $G$ .
- (2)  $S(G, \Gamma) = C(G, \Gamma)$ .

*Proof.*  $1 \Rightarrow 2$ . Let  $\Gamma \subset G$ ,  $G' \subset \text{Hol}(U)$  and let  $\phi: G' \rightarrow G$  be an isomorphism with the algebraic extension  $\tilde{\phi}: A(G') \rightarrow A(G)$ . Put  $\Gamma' = \phi(\Gamma) \subset G$ . Since  $\Gamma$  is rigid in  $G$ , the isomorphism  $\phi: \Gamma \rightarrow \Gamma'$  has an extension  $\psi: G \rightarrow G'$ . It follows from theorem 3.1 that there exists an algebraic extension  $\tilde{\psi}: A(G) \rightarrow A(G')$ . Hence,  $\tilde{\phi}^{-1} \circ \tilde{\psi}: A(G) \rightarrow A(G')$  is an algebraic isomorphism such that  $(\tilde{\phi}^{-1} \circ \tilde{\psi})|_\Gamma \equiv \text{Id}$  and by corollary 3.5,  $G' = \phi^{-1}\psi(G) \in C(G, \Gamma)$ .

$2 \Rightarrow 1$ . Let  $\phi: \Gamma' \rightarrow \Gamma$  be an isomorphism with some algebraic extension  $\phi^*: \text{Hol}(U) \rightarrow \text{Hol}(U)$ . Then  $\phi^*(G) \supset \phi(\Gamma') = \Gamma$  and  $\phi^*(G) \simeq G$ . Hence,  $\phi^*(G) = zGz^{-1}$  for some  $z \in \text{Hol}(U)$  commuting with  $\Gamma$ . Let  $\hat{\phi}(g) = z^{-1}\phi^*(g)z$ , for all  $g \in G$ . Then  $\hat{\phi}|_{\Gamma'} \equiv \phi^*|_{\Gamma'} \equiv \phi$  and  $\hat{\phi}(G) = G$ . Thus, an arbitrary isomorphism  $\phi: \Gamma' \rightarrow \Gamma$  has an extension  $\hat{\phi}: G \rightarrow G$ .

*Remark.* An example of a rigid lattice  $\Gamma$  in a solvable Lie group  $G$  with nontrivial family  $C(G, \Gamma)$  is unknown to the author. It is very likely that  $\Gamma$  is rigid in  $G$  if and only if the family  $S(G, \Gamma)$  is trivial.

The following corollaries of the rigidity criterion will be useful to us in future:

### COROLLARY 3.7

Assume that  $\Gamma$  is a Zariski-dense lattice in  $G$  or the unipotent hull  $U$  of  $G$  is abelian. Then  $\Gamma$  is rigid in  $G \Leftrightarrow$  the family  $S(G, \Gamma)$  is trivial.

*Proof.* If  $\Gamma$  is Zariski-dense, then  $Z_U(T_\Gamma) = Z_U(A(T_\Gamma)) = Z_U(A(T_G))$ , and hence  $mm_*^{-1}$  commutes with  $A(G)$  for all  $m \in Z_U(T_\Gamma)$ . If  $U$  is an abelian group, then  $\text{Int}(U) = 1$  and  $mm_*^{-1}Gm_*m^{-1} = mGm^{-1} = G$ , for all  $m \in Z_U(T_\Gamma)$ , as well. Hence, in both these cases the family  $C(G, \Gamma)$  is trivial.

It is not difficult now to deduce the rigidity theorem 2.4 of Saito. In fact, let  $G$  be a simply connected solvable Lie group of (R)-type with a lattice  $\Gamma \subset G$ . Then by proposition 1.11, the algebraic hull  $A(G) \subset \text{Hol}(U)$  is simply connected, and hence any lattice is Zariski-dense in  $G$ . Assume that  $\Gamma \subset G' \subset \text{Hol}(U)$  and  $G' \simeq G$ . Then the nilradical of  $G'$  contains  $A[\Gamma, \Gamma] = [A(\Gamma), A(\Gamma)] = [A(G), A(G)]$ . But since the quotient group  $A(G)/[A(G), A(G)]$  is simply connected and abelian, it follows that  $G$  is the unique connected subgroup in  $A(G)$ , containing  $\Gamma[A(G), A(G)]$  as a uniform subgroup. Hence,  $G' = G$  and by corollary 3.7,  $\Gamma$  is rigid in  $G$ .

### COROLLARY 3.8

*Assume that there are at least two elements of the family  $S(G, \Gamma)$  inside  $A(G)$ . Then the lattice  $\Gamma$  is nonrigid in  $G$ .*

*Proof.* It suffices to prove that  $G$  is the unique element of  $C(G, \Gamma)$  inside  $A(G)$ . So assume that  $G_m = mm_*^{-1}Gm_*m^{-1} = m_*^{-1}Gm_* \subset A(G)$  for some  $m \in Z_U(T_\Gamma) \subset U$ . Then  $A(G_m) \subset A(G)$ , and hence the unipotent element  $m_* \in \text{Aut}(U)$  normalizes  $A(T_G)$ . This is possible only if  $m_*$  commutes with  $A(T_G)$ . Hence, for every  $a \in A(T_G)$ ,  $n \in U$ , we have  $m_*^{-1}anm_* = am_*^{-1}nm_* = am^{-1}nm$ . This means that  $g^{-1}m_*^{-1}gm_* \in [U, U] \subset G$ , for all  $g \in G$ , and, in particular,  $G_m = G$ .

### COROLLARY 3.9

*Let  $\bar{\psi}(G) = G$  for all canonical isomorphisms  $\bar{\psi}: \text{Hol}(U) \rightarrow \text{Hol}(U)$  such that  $\bar{\psi}(\Gamma) \subset G$ . Then  $\Gamma$  is rigid in  $G$ .*

*Proof.* By corollaries 3.4 and 3.5, for any algebraic extension  $\phi^*: \text{Hol}(U) \rightarrow \text{Hol}(U)$  of an isomorphism  $\phi: \Gamma \rightarrow \Gamma'$  there exist  $u \in U$ ,  $m \in Z_U(T_\Gamma)$  and a canonical isomorphism  $\psi: \text{Hol}(U) \rightarrow \text{Hol}(U)$  such that  $\phi^*(g) = mm_*^{-1}u\psi(g)u^{-1}m_*m^{-1}$  for all  $g \in \text{Hol}(U)$ . But  $mm_*^{-1}u = umm_*^{-1}$  and the formula  $\bar{\psi}(g) = m_*^{-1}\psi(g)m_*$  for all  $g \in \text{Hol}(U)$  determines a canonical isomorphism  $\bar{\psi}$ . Consequently,  $\phi^*(g) = um\bar{\psi}(g)m^{-1}u^{-1}$  for all  $g \in \text{Hol}(U)$ . Obviously,  $\phi^*(\Gamma) \subset G \Leftrightarrow \bar{\psi}(\Gamma) \subset G$  and  $\phi^*(G) = G \Leftrightarrow \bar{\psi}(G) = G$ , since  $U$  normalizes  $G$ . In view of theorem 3.6, this completes the proof.

## 4. Rigidity and type (R)

Up to the present time, all examples of rigid lattices known to us are those in solvable Lie groups of (R)-type. The natural question arises: is the class (R) the maximal class of solvable Lie groups with rigid lattices. Here we show that the answer is negative: we construct a new example of a rigid lattice in a solvable Lie group not even of (E)-type.

**Example 4.1** This is a construction from Auslander's paper [1] which presents the first nontrivial example of a solvable Lie group of (A)-type. Let us view the matrix

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$



as an operator on a vector space  $V^4$  with some given initial basis. It is easy to verify that eigenvalues of  $A$  are as follows:  $\lambda_{1,2} = e^{\pm t_0}$ ,  $\lambda_{3,4} = e^{\pm 2\pi i \theta_0}$ , where  $t_0, \theta_0 \in \mathbf{R}$ ,  $t_0 > 1$  and  $\theta_0$  is irrational. Let  $V^4 = V' \times V''$  be the decomposition of  $V^4$  into invariant planes  $V'$  and  $V''$  corresponding to the pairs of eigenvalues  $\lambda_{1,2}$  and  $\lambda_{3,4}$  respectively. Let

$$H(t) = \begin{pmatrix} e^t & 0 & 0 & 0 \\ 0 & e^{-t} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad O(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos(2\pi\theta) & -\sin(2\pi\theta) \\ 0 & 0 & \sin(2\pi\theta) & \cos(2\pi\theta) \end{pmatrix}$$

in the basis of  $V^4$  formed by the eigenvectors of  $A$ . Then  $H(\mathbf{R})$  is the one-parameter subgroup of  $\text{Aut}(V^4)$  of hyperbolic actions on  $V'$  and  $O(\mathbf{R})$  is the circle in  $\text{Aut}(V^4)$  of orthogonal rotations on  $V''$ .

Let  $D(t) = H(tt_0) \times O(t\theta_0)$ ,  $t \in \mathbf{R}$ . Then  $D(\mathbf{R})$  is a one-parameter subgroup of the cylinder  $H(\mathbf{R}) \times O(\mathbf{R}) \subset \text{Aut}(V^4)$ . The action of the element  $D(1)$  in the initial basis of  $V^4$  is represented by the integer matrix  $A$ . Hence, we may form the Lie group  $G = D(\mathbf{R}) \cdot V^4$  with the lattice  $\Gamma = D(\mathbf{Z}) \cdot \mathbf{Z}^4$ , where  $\mathbf{Z}^4$  is the integer lattice in  $V^4$  relative to the initial basis of  $V^4$  and  $D(\mathbf{Z})$  is the cyclic group generated by the element  $D(1) = A$ .

Clearly,  $G$  is a simply connected solvable Lie group of  $(A) \setminus (E)$ -type. Note that  $G$  has the subgroup  $D(\mathbf{R}) \cdot V''$  of  $(I)$ -type. Let us prove that  $\Gamma$  is rigid in  $G$ . The group  $G$  is splittable and by proposition 1.7, its semisimple splitting has the form  $G_S = T_G \cdot U$ , where  $U \simeq V^4 \times \mathbf{R}$  and  $T_G = D^*(\mathbf{R})$  is the image of  $D(\mathbf{R}) \subset \text{Aut}(V^4)$  under the injection  $*: \text{Aut}(V^4) \hookrightarrow \text{Aut}(V^4 \times \mathbf{R})$ . Besides,  $A(T_G) = H^*(\mathbf{R}) \times O^*(\mathbf{R})$  and  $A(T_\Gamma) = A(D^*(\mathbf{Z})) = H^*(\mathbf{R}) \times O^*(\mathbf{R})$ , since  $\theta_0$  is irrational. Therefore, the lattice  $\Gamma$  is Zariski-dense in  $G$ .

Assume that  $\Gamma$  is a lattice in a simply connected solvable Lie group  $G' \subset \text{Hol}(U)$  and  $G' \simeq G$ . Then by proposition 1.12, the nilradical  $N'$  of  $G'$  has the form  $N' = G' \cap U$  and by Mostow's theorem 1.8, it contains the lattice  $N' \cap \Gamma = \Gamma \cap U$ . Hence,  $N' = A(\Gamma \cap U) = V^4$ . Since  $\Gamma$  must be Zariski-dense in  $G'$ , it follows that  $A(G') = A(\Gamma) = A(G)$  and we may assume that  $T_{G'} \subset A(T_G)$ .

Let us prove that  $T_{G'} = T_G$ . In fact,  $T_{G'}$  is a one-parameter subgroup of the cylinder  $H^*(\mathbf{R}) \times O^*(\mathbf{R})$  containing the cyclic group  $D^*(\mathbf{Z})$ . But any such group has the form  $\{D_n^*(t) = H^*(tt_0) \times O^*(t(\theta_0 + n))\}$ ,  $t \in \mathbf{R}$  for some  $n \in \mathbf{Z}$ . The eigenvalues of the generator of the one-parameter group  $D_n^*(\mathbf{R})$  are as follows:  $\lambda_{1,2} = \pm t_0$ ,  $\lambda_{3,4} = \pm 2\pi i(\theta_0 + n)$ . Hence, the action of  $D_n^*(\mathbf{R})$  on  $V^4$  is non-isomorphic to that of  $D^*(\mathbf{R})$  provided  $n \neq 0$ . This shows that  $T_{G'} = T_G = D^*(\mathbf{R})$ .

Finally, since  $\Gamma \subset G' \subset G_S$  and the quotient abelian group  $G_S/V^4$  is simply connected, it follows that  $G$  is the unique connected subgroup in  $G_S$  having  $\Gamma V^4$  as a uniform subgroup. This proves that  $G' = G$  and by corollary 3.5, the lattice  $\Gamma$  is rigid in  $G$ .

## 5. Rigidity and type (A)

We have constructed in §4 an example of a rigid lattice in a splittable solvable Lie group of  $(A)$ -type. Now we prove that for splittable Lie groups the class  $(A)$  is the maximal class of solvable groups having a rigid lattice. Note that in §8 we construct an example of a rigid lattice in a (nonsplittable) solvable Lie group of  $(I)$ -type.

But first of all let us prove that the class (A) is the maximal opposite class to the class (I). To be more precise, for every solvable Lie group  $G$  with a lattice  $\Gamma \subset G$  we construct the decomposition of  $G$  and  $\Gamma$  into their maximal parts of (A) and (I)-types.

**Theorem 5.1** *Let  $G$  be a simply connected solvable Lie group with a lattice  $\Gamma \subset G$ . Then there exist connected normal subgroups  $G_I, G_A \subset G$  such that*

- (1)  $G = G_I G_A$  and  $G_I \cap G_A = N$ , where  $N$  is the nilradical in  $G$ ,
- (2)  $G_I$  is of (I)-type,
- (3)  $G_A$  is of (A)-type, and
- (4)  $\Gamma = (\Gamma \cap G_I)(\Gamma \cap G_A)$ .

*Proof.* Let  $G_S = T_G \cdot U$  be the semisimple splitting of  $G$ . Then by Mostow's theorem 1.8, the closed abelian subgroup  $T_G \subset \text{Aut}(U)$  has a lattice  $T_\Gamma = p(\Gamma)$ , where  $p: G_S = T_G U \rightarrow T_G$  is the natural projection. Hence, there exists a decomposition  $T_G = T_I^* \times T_A^*$  of  $T_G$  into its maximal compact subgroup  $T_I^*$  and some simply connected subgroup  $T_A^*$  such that  $p(\Gamma) = (p(\Gamma) \cap T_I^*)(p(\Gamma) \cap T_A^*)$ . Let  $G_I = (T_I^* U \cap G)_0$  and  $G_A = (T_A^* U \cap G)_0$ . It is an easy verification that  $G_I$  and  $G_A$  are connected normal subgroups of  $G$  having the types (I) and (A) respectively.

Since  $N \subset G_I \cap G_A$  and  $p(G) = p(G_I)p(G_A) = T_I^* T_A^*$ , it follows from the dimensional argument that  $G = G_I G_A$ . It suffices to prove that  $N = G_I \cap G_A$  and  $\Gamma = (\Gamma \cap G_I)(\Gamma \cap G_A)$ .

We are able to prove the stronger equalities:  $G_I \cap U = G \cap U$  and  $G_A \cap U = N$ . Let us consider the vector space  $G/N$  with the discrete subgroup  $C = (G \cap U)/N$ . If  $A$  is a linear hull of  $C$  then  $A/C \simeq T_I^*$ . Put  $C' = (G_I \cap U)/N$  and  $A' = G_I/N$ . Then  $A'$  is a linear hull of  $C' = C \cap A'$  and  $A'/C' \simeq T_I^*$ . Hence,  $A = A'$  and  $G_I \cap U = G \cap U$ .

Now let us consider the bundle  $G_A \rightarrow T_A^*$  with the fibre  $G_A \cap U$ . Since  $G_A$  is connected and  $T_A^*$  is simply connected, it follows that the fiber  $G_A \cap U$  is connected, and hence  $G_A \cap U = N$ .

Finally, let  $\gamma \in \Gamma$ . Since  $p(\Gamma) = (p(\Gamma) \cap T_I^*)(p(\Gamma) \cap T_A^*)$ , there exist  $\gamma_1, \gamma_2 \in \Gamma$  such that  $p(\gamma_1) \in T_I^*, p(\gamma_2) \in T_A^*$  and  $p(\gamma) = p(\gamma_1 \gamma_2)$ . But the restrictions  $p: G_I \rightarrow T_I^*$  and  $p: G_A \rightarrow T_A^*$  are epimorphisms and we may assume that  $\gamma_1 \in G_I, \gamma_2 \in G_A$ . Hence,  $\gamma = n\gamma_1 \gamma_2$ , where  $n \in U$ . But then  $n \in \Gamma \cap U \subset G \cap U \subset G_I$  and this completes the proof.

*Remark.* It may be noted that the maximal subgroup  $G_I$  of (I)-type is uniquely determined by  $G$ , while a maximal subgroup  $G_A$  of (A)-type need not be unique even for a fixed  $\Gamma \subset G$ .

Now we are able to prove a very simple necessary condition for a splittable Lie group  $G$  to have a rigid lattice.

## PROPOSITION 5.2

*Let  $G$  be a splittable simply connected solvable Lie group with a rigid lattice  $\Gamma \subset G$ . Then  $G$  is of (A)-type. In particular, every lattice in a splittable solvable Lie group of (I) \setminus (N)-type is nonrigid.*

*Proof.* Let  $G = T \cdot N$  be a splitting of  $G$  into its nilradical  $N$  and an abelian subgroup  $T$  of semisimple elements. Since the groups  $G_I, G_A$  from the decomposition 5.1 contain

$N$ , it follows that these groups split as well:  $G_I = T_I \cdot N$ ,  $G_A = T_A \cdot N$ . It is obvious that  $p(T_I) = T_I^*$ ,  $p(T_A) = T_A^*$ , where  $p: G_S = T_G \cdot U \rightarrow T_G$  is the natural projection,  $T_I^*$  is the maximal compact subgroup of  $T_G$  and  $T_A^*$  is simply connected. Besides, by proposition 1.7, the unipotent hull  $U$  splits into the product of the nilradical  $N$  with the "antidiagonal"  $\Delta \subset T_G \times T$ . Hence,  $U = N \times \Delta_I \times \Delta_A$ , where  $\Delta_I, \Delta_A$  are the "antidiagonals" in  $T_I^* \times T_I$ ,  $T_A^* \times T_A$  respectively.

Let  $\alpha: G \rightarrow \Delta_I$  be the composition of the diffeomorphism  $\pi: G \rightarrow U$  along  $T_G$  with the projection  $U = \Delta_I \times \Delta_A \times N \rightarrow \Delta_I$  along  $\Delta_A \times N$ . Then  $\alpha$  is easily seen to be an epimorphism of groups with the kernel  $G_A$ . Note that  $T_I^* = T_I / (T_I \cap \Delta_I) \simeq \Delta_I / (T_I \cap \Delta_I)$  and since  $T_I^*$  is compact,  $\Delta_I \cap T_I = \Delta_I \cap G$  is a lattice in  $\Delta_I$  and  $\Gamma \cap U$  is of finite index in  $\Gamma \cap G_I$ . Since  $\Gamma \cap U$  is a lattice in  $G_I$  and  $\Delta_I = \alpha(G_I)$ , it follows that  $\alpha(\Gamma \cap U)$  is a lattice in  $\Delta_I$ .

Thus we have three lattices in  $\Delta_I: \Delta_I \cap G$ ,  $\alpha(\Gamma \cap U)$ , and  $\alpha(\Gamma)$ . Since  $G_S = (T_G \cdot N) \times \Delta_I \times \Delta_A$ , for any  $\sigma \in \text{Aut}(\Delta_I)$  there exists the trivial extension  $\hat{\sigma} \in \text{Aut}(G_S)$ . It suffices to find  $\sigma \in \text{Aut}(\Delta_I)$  such that  $\hat{\sigma}(\Gamma) \subset G$  and  $\hat{\sigma}(G) \neq G$ . In this case  $G$  and  $\hat{\sigma}^{-1}(G)$  will be two different isomorphic groups inside  $A(G)$  having  $\Gamma$  as a lattice and by corollary 3.8, this means that  $\Gamma$  is nonrigid in  $G$ .

It may be easily proved that  $\hat{\sigma}(g)g^{-1} = \sigma(\alpha(g))\alpha(g)^{-1}$  for every  $g \in G$ . Hence, if  $\sigma(z)z^{-1} \in G$  for every  $z \in \alpha(\Gamma)$  then  $\hat{\sigma}(\gamma)\gamma^{-1} \in G$  for every  $\gamma \in \Gamma$  and so  $\hat{\sigma}(\Gamma) \subset G$ . On the other hand, if  $\sigma \neq Id$  then  $\hat{\sigma}(G) \neq G$  (Indeed, if  $\hat{\sigma}(G) = G$  then  $\hat{\sigma}(g)g^{-1} \in \Delta_I \cap G$ . But  $\Delta_I \cap G$  is discrete, while  $G$  is connected!).

Let  $\beta: \alpha(\Gamma) \rightarrow \Delta_I \cap G$  be an arbitrary expanding homomorphism of lattices:  $|\beta(z)| > |z|$  for every  $z \in \alpha(\Gamma)$ . Now if  $\hat{\beta}: \Delta_I \rightarrow \Delta_I$  is the linear extension of  $\beta$  and  $\sigma(z) = \hat{\beta}(z)z$  then  $\sigma \in \text{Aut}(\Delta_I)$ ,  $\sigma \neq Id$  and  $\sigma(z)z^{-1} = \beta(z) \in \Delta_I \cap G$  for any  $z \in \alpha(\Gamma)$ . Hence,  $\hat{\sigma}(\Gamma) \subset G$  and  $\hat{\sigma}(G) \neq G$ . This proves nonrigidity of  $\Gamma$ , provided  $G \neq G_A$ .

*Remark.* Note that we used the same method to prove nonrigidity of the virtually abelian lattice  $S(\frac{1}{2}\pi\mathbf{Z}) \cdot \mathbf{Z}^2$  in the splittable Lie group  $G = S(\mathbf{R}) \cdot \mathbf{R}^2$  of (I)-type in example 2.8.

## 6. Rigidity and the Zariski-density

Note that up to now, every example of a rigid lattice in a solvable group has also been Zariski-dense. Conversely, the non-rigid lattices in examples 2.6–2.9 were not Zariski-dense. However, these two properties for a lattice in a solvable Lie group are quite different, as the following two examples show. First we construct example 6.1 of a lattice which is rigid but not Zariski-dense, then we construct example 6.2 of a lattice which is Zariski-dense but not rigid. In both examples  $G$  is a splittable solvable Lie group of (A)-type.

*Example 6.1.* Consider the Lie group  $G_1 = D_1(\mathbf{R}) \cdot V_1^4$  from example 4.1 with the lattice  $\Gamma_1 = D_1(\mathbf{Z}) \cdot \mathbf{Z}_1^4$ , where the action of  $D_1(1)$  on the vector space  $V_1^4$  is represented by the integer matrix  $A$ . Recall that we have a decomposition  $V_1^4 = V_1' \times V_1''$  of  $V_1^4$  into invariant planes, and an injection  $D_1(\mathbf{R}) \subset H_1(\mathbf{R}) \times O_1(\mathbf{R})$ , where  $H_1(\mathbf{R})$  acts on  $V_1'$  by hyperbolic rotations and  $O_1(\mathbf{R})$  acts on  $V_1''$  by orthogonal rotations. To be more precise,  $D_1(t) = H_1(tt_0) \times O_1(t\theta_0)$  with some  $t_0 > 1$  and irrational  $\theta_0$ .

We have proved that  $\Gamma_1$  is a rigid Zariski-dense lattice in the solvable Lie group

$G_1$  of (A)-type. To construct the desired example, consider the copy  $V_2^4 = V_2' \times V_2''$  and the corresponding cylinder  $H_2(\mathbf{R}) \times O_2(\mathbf{R}) \subset \text{Aut}(V_2^4)$ . Put  $G_2 = D_2(\mathbf{R}) \cdot V_2^4$ , where  $D_2(t) = H_2(tt_0) \times O_2(t(\theta_0 + 1/2))$  for all  $t \in \mathbf{R}$ , and  $\Gamma_2 = D_2(2\mathbf{Z}) \cdot \mathbf{Z}_2^4$ . Note that  $\Gamma_2$  is isomorphic to a proper subgroup in  $\Gamma_1$  and  $G_2$  is nonisomorphic to  $G_1$ .

Now consider the group  $G = D(\mathbf{Z}) \cdot (V_1^4 \times V_2^4)$ , where  $D(t) = D_1(t) \times D_2(2t)$ , with the lattice  $\Gamma = D(\mathbf{Z}) \cdot (\mathbf{Z}_1^4 \times \mathbf{Z}_2^4)$ . The group  $G$  is a splittable simply connected solvable Lie group of (A)-type and by proposition 1.7, it follows that  $G_S = T_G \cdot U$ , where  $T_G = D^*(\mathbf{R})$ ,  $U = V_1^4 \times V_2^4 \times \Delta$ ,  $D^*(\mathbf{R})$  is the image of  $D(\mathbf{R})$  under the injection  $^* : \text{Aut}(V_1^4 \times V_2^4) \hookrightarrow \text{Aut}(V_1^4 \times V_2^4 \times \Delta)$  and  $\Delta$  is the “antidiagonal” in  $D^*(\mathbf{R}) \times D(\mathbf{R})$ .

It is a straightforward verification that  $A(\Gamma) = (H^*(\mathbf{R}) \times O^*(\mathbf{R})) \cdot U$ , where  $H^*(\mathbf{R}) \subset H_1^*(\mathbf{R}) \times H_2^*(\mathbf{R})$  and  $O^*(\mathbf{R})$  is a one-dimensional subtorus in the torus  $O_1^*(\mathbf{R}) \times O_2^*(\mathbf{R})$  (to be more precise,  $H^*(t) = H_1^*(t) \times H_2^*(2t)$ ,  $O^*(t) = O_1^*(t) \times O_2^*(2t)$ ). On the other hand,  $A(G) = (H^*(\mathbf{R}) \times O_1^*(\mathbf{R}) \times O_2^*(\mathbf{R})) \cdot U$ , since the group  $\{O_1^*(t\theta_0) \times O_2^*(2t\theta_0 + t), t \in \mathbf{R}\}$  is dense in the torus  $O_1^*(\mathbf{R}) \times O_2^*(\mathbf{R})$  due to the irrationality of  $\theta_0$ . Hence, the lattice  $\Gamma$  is not Zariski-dense in  $G$ .

Let us prove that  $\Gamma$  is rigid in  $G$ . Assume that  $\Gamma \subset \tilde{G} \subset \text{Hol}(U)$  and  $\tilde{G} \simeq G$ . This means, in particular, that  $\tilde{G}$  is splittable, and as always, it may be easily proved that  $\tilde{G} = \tilde{D}(\mathbf{R}) \cdot (V_1^4 \times V_2^4)$ , where  $D(\mathbf{Z}) \subset \tilde{D}(\mathbf{R})$ . It suffices to prove that  $\tilde{D}(\mathbf{R}) = D(\mathbf{R})$ . But first we show that  $\tilde{D}(\mathbf{R}) \subset H_1(\mathbf{R}) \times H_2(\mathbf{R}) \times O_1(\mathbf{R}) \times O_2(\mathbf{R})$ .

In fact, since  $\Gamma$  is a lattice in  $\tilde{G}$ , it follows from theorem 1.9 that  $A(\tilde{G}) = CA(\Gamma)$ , where  $C$  is a torus in  $\text{Aut}(U)$  which commutes with  $A(T_\Gamma) = H^*(\mathbf{R}) \times O^*(\mathbf{R})$ . Hence, the torus  $C$  must keep invariant the weight spaces of the action of  $H^*(\mathbf{R})$  on  $U = V_1^4 \times V_2^4 \times \Delta$ . The generator of  $H^*(\mathbf{R})$  has the following eigenvalues:  $\lambda_1 = 0$  with multiplicity 5 and  $\lambda_{2,3} = \pm 1$ ,  $\lambda_{4,5} = \pm 2$ . Clearly, the torus  $C$  may act nontrivially only on the weight space  $V_1'' \times V_2'' \times \Delta$  of the eigenvalue  $\lambda_1 = 0$ . On the other hand, the generator of  $O^*(\mathbf{R})$  has the following eigenvalues on  $V_1'' \times V_2'' \times \Delta$ :  $\lambda_{1,2} = \pm 2\pi i$ ,  $\lambda_{3,4} = \pm 4\pi i$ ,  $\lambda_5 = 0$ . Clearly, the torus  $C$  acts trivially on  $\Delta$  and keeps invariant the subspaces  $V_1''$  and  $V_2''$ . Hence,  $C \subset O_1^*(\mathbf{R}) \times O_2^*(\mathbf{R})$  and  $\tilde{D}(\mathbf{R}) \subset H_1(\mathbf{R}) \times H_2(\mathbf{R}) \times O_1(\mathbf{R}) \times O_2(\mathbf{R})$ .

Since  $D(\mathbf{Z}) \subset \tilde{D}(\mathbf{R})$ , it is not difficult to verify that  $\tilde{D}(\mathbf{R})$  has the form  $\{D_{n,m}(t) = H_1(tt_0) \times H_2(2tt_0) \times O_1(t(\theta_0 + n)) \times O_2(t(2\theta_0 + m)), t \in \mathbf{R}\}$  for some  $n, m \in \mathbf{Z}$ . But the generator of  $D_{n,m}(\mathbf{R})$  has the following eigenvalues on  $V_1^4 \times V_2^4$ :  $\lambda_{1,2} = \pm t_0$ ,  $\lambda_{3,4} = \pm 2t_0$ ,  $\lambda_{5,6} = \pm 2\pi i(\theta_0 + n)$ ,  $\lambda_{7,8} = \pm 2\pi i(2\theta_0 + m)$ . Since  $\tilde{G} \simeq G$ , it follows that  $\tilde{D}(\mathbf{R}) = \tilde{D}_{0,1}(\mathbf{R}) = D(\mathbf{R})$  and  $\tilde{G} = G$ . In view of corollary 3.7 this proves that  $\Gamma$  is rigid in  $G$ , but not Zariski-dense in  $G$ .

*Example 6.2.* The construction of a nonrigid Zariski-dense lattice is much easier. Again let  $G_1 = D_1(\mathbf{R}) \cdot V_1^4$ , where  $D_1(t) = H_1(tt_0) \times O_1(t\theta_0)$ , and  $\Gamma_1 = D_1(\mathbf{Z}) \cdot \mathbf{Z}_1^4$ . Similarly, put  $G_2 = D_2(\mathbf{R}) \cdot V_2^4$ ,  $\Gamma_2 = D_2(\mathbf{Z}) \cdot \mathbf{Z}_2^4$ , where  $D_2(t) = H_2(tt_0) \times O_2(t(\theta_0 + 1))$ . Since  $O_2((\theta_0 + 1)\mathbf{Z}) = O_2(\theta_0\mathbf{Z})$ , it follows that  $\Gamma_1 \simeq \Gamma_2$ . At the same time, the groups  $G_1$  and  $G_2$  are nonisomorphic (this follows by considering the eigenvalues of the generators of  $D_1(\mathbf{R})$  and  $D_2(\mathbf{R})$ ). The lattices  $\Gamma_1$  and  $\Gamma_2$  are Zariski-dense (and rigid) in  $G_1$  and  $G_2$  respectively. Put  $G = G_1 \times G_2$  and  $\Gamma = \Gamma_1 \times \Gamma_2$ . Then the automorphism of  $\Gamma$  which permutes  $\Gamma_1$  and  $\Gamma_2$  cannot be extended to an automorphism of  $G$ , since  $G_1$  and  $G_2$  are nonisomorphic. Consequently,  $\Gamma$  is not weakly rigid in  $G$ .

## 7. Rigidity and deformations

Here we consider the problem of extending a deformation of a lattice  $\Gamma \in G$  to an automorphism of  $G$ .

### DEFINITION 7.1

Let  $R(\Gamma, G)$  be the topological space, with the topology of pointwise convergence, of all isomorphisms  $\phi: \Gamma \rightarrow \phi(\Gamma) \subset G$  of  $\Gamma$  onto another lattice in  $G$ , and let  $R(\Gamma, G)_0$  be the connected component of the identity map  $Id \in R(\Gamma, G)$ . Then any element  $\phi \in R(\Gamma, G)_0$  is said to be a *deformation* of  $\Gamma$  in  $G$ .

It is well known that the space  $R(\Gamma, G)_0$  is arcwise connected, i.e. any deformation  $\phi \in R(\Gamma, G)_0$  may be connected with the identity  $Id \in R(\Gamma, G)$  by a continuous curve. Clearly, any rigid lattice is rigid under deformations. The converse statement is false. This means that an extension may exist for every deformation  $\phi \in R(\Gamma, G)_0$  but not for every isomorphism  $\phi \in R(\Gamma, G)$ .

In fact, let us recall the example 2.8 of the group  $G = S(\mathbf{R}) \cdot \mathbf{R}^2$ , where the one-parameter group  $S(\mathbf{R})$  acts on the plane  $\mathbf{R}^2$  by orthogonal rotations with the kernel  $S(2\pi\mathbf{Z})$ . We proved that the lattice  $\Gamma = S(\frac{1}{2}\pi\mathbf{Z}) \cdot \mathbf{Z}^2$  is weakly rigid and nonrigid in  $G$ . Now we may prove that this lattice is rigid under deformations using the following fundamental result due to H.-C. Wang:

**Theorem 7.1** [22]. *Let  $\Gamma$  be a lattice in a simply connected solvable Lie group  $G$ , and let  $N$  be the nilradical in  $G$ . Then  $\phi(\gamma)N = \gamma N$  for all  $\gamma \in \Gamma$  and any deformation  $\phi$  of  $\Gamma$ .*

Let  $\phi \in R(\Gamma, G)$  be a deformation of  $\Gamma = S(\pi/2\mathbf{Z}) \cdot \mathbf{Z}^2$  in the group  $G = S(\mathbf{R}) \cdot \mathbf{R}^2$ . Then  $\phi(s(\pi/2)) = s(\pi/2)v$  for some  $v \in \mathbf{R}^2$ . Hence, there exists an element  $m \in \mathbf{R}^2$  such that  $\phi(s(\pi/2)) = ms(\pi/2)m^{-1}$ . Note that  $\phi(\mathbf{Z}^2) = \phi([\Gamma, \Gamma]) = [\phi(\Gamma), \phi(\Gamma)]$  is a lattice in the nilradical  $\mathbf{R}^2$  of  $G$ ; consequently, the restriction  $\phi: \mathbf{Z}^2 \rightarrow \mathbf{R}^2$  determines an element  $\phi_* \in \text{Aut}(\mathbf{R}^2)$ .

Since the actions of  $s(\pi/2)$  and  $\phi(s(\pi/2)) = ms(\pi/2)m^{-1}$  on  $\mathbf{R}^2$  are represented by the matrix

$$a^* = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

it follows that  $\phi_*$  commutes with  $a^*$ . But the centralizer of  $a^*$  in  $\text{Aut}(\mathbf{R}^2)$  coincides with the centralizer of  $SO(2)$ , and hence  $\phi_*$  commutes with the action of  $S(\mathbf{R})$  on  $\mathbf{R}^2$ . Thus, the rule  $\hat{\phi}(an) = mam^{-1}\phi_*(n)$ , for all  $a \in S(\mathbf{R})$ ,  $n \in \mathbf{R}^2$ , determines the required extension  $\hat{\phi}: G \rightarrow G$ .

Now let us give the following simple sufficient condition for a lattice to be rigid under deformations:

### PROPOSITION 7.2

*Let the family  $S(G, \Gamma)$  of all subgroups  $G' \subset \text{Hol}(U)$  such that  $\Gamma \subset G'$  and  $G' \simeq G$  be at most countable. Then any deformation  $\phi \in R(\Gamma, G)_0$  has an extension  $\hat{\phi}: G \rightarrow G$ .*

*Proof.* The idea is as follows. By theorem 3.1, any continuous family  $\phi_t \in R(\Gamma, G)$  determines a continuous family of algebraic isomorphisms  $\hat{\phi}_t: A(\Gamma) \rightarrow A(\Gamma_t)$ , where

$\Gamma_t = \phi_t(\Gamma)$ ,  $t \in [0, 1]$ . It follows from proposition 3.4 that for every  $t \in [0, 1]$  there exists an element  $m_t \in U$  such that  $\hat{\phi}_t(an) = (mm_t^{-1} \phi_*^{-1})a(\phi_*^{-1} m_t m^{-1}) \phi_*(n)$  for all  $a \in A(T_\Gamma)$ ,  $n \in U$ . It may be proved as well that without loss of generality we may assume that the family  $\{m_t, t \in [0, 1]\}$  is continuous. Then if we define  $\hat{\phi}_t(an)$  as above for all  $a \in \text{Aut}(U)$ ,  $n \in U$ , we obtain a continuous family  $\hat{\phi}_t: \text{Hol}(U) \rightarrow \text{Hol}(U)$  of algebraic isomorphisms. This gives us a continuous family  $\{G_t = \hat{\phi}_t(G)\}$  of Lie groups  $G_t$  with lattices  $\Gamma_t$ . Since  $\Gamma_t \subset G$ , this provides us with a continuous subfamily  $\{\hat{\phi}_t^{-1}(G), t \in [0, 1]\}$  of the family  $S(G, \Gamma)$ . But since  $S(G, \Gamma)$  is at most countable, it follows that  $G_t = G$  for all  $t \in [0, 1]$ . In particular, the automorphism  $\hat{\phi}_1: G \rightarrow G$  is an extension of the deformation  $\phi_1: \Gamma \rightarrow \Gamma_1$ .

### COROLLARY 7.3

*Any Zariski-dense lattice is rigid under deformations.*

*Proof.* Since  $\Gamma$  is Zariski-dense in  $G$ , it follows that  $A(G) = A(\Gamma) = A(G')$  for any  $G' \in S(G, \Gamma)$ . Note that the commutator  $[\Gamma, \Gamma]$  of the lattice  $\Gamma \subset G'$  lies in the nilradical  $N' = (G' \cap U)_0$  of  $G'$ . Hence,  $[A(G), A(G)] = [A(\Gamma), A(\Gamma)] = A([\Gamma, \Gamma]) \subset G'$ . But the abelian group  $A(G)/[A(G), A(G)]$  contains at most a countable set of connected subgroups having  $\Gamma/[\Gamma, \Gamma]$  as a lattice. This means that the family  $S(G, \Gamma)$  is at most countable.

*Remark.* This corollary may be deduced as well from Mostow's results in [13].

In particular, the Zariski-dense lattice  $\Gamma$  in example 6.2 is rigid under deformations, not being weakly rigid.

## 8. Rigidity and type (I)

We proved in §5 that there are no rigid lattices in a splittable solvable Lie group of  $(I) \setminus (N)$ -type. However, a nonsplittable group of  $(I)$ -type may contain a rigid lattice. To construct this example let us first recall one important result, conjectured by Grunewald and Segal [9] and proved by Bryant and Groves [7].

Let  $\mathfrak{n}$  be a nilpotent Lie algebra and let  $V = \mathfrak{n}/[\mathfrak{n}, \mathfrak{n}]$  be an "abelianization" of  $\mathfrak{n}$ . Then we have a natural map  $*$ :  $\text{Aut}(\mathfrak{n}) \rightarrow \text{Aut}(V) = GL(V)$ . Clearly,  $\text{Ker} *$  lies in the unipotent radical of  $\text{Aut}(\mathfrak{n})$ . Put  $\text{Aut}^*(\mathfrak{n}) = *(\text{Aut}(\mathfrak{n})) \subset GL(V)$ . The following theorem gives a description of groups  $\text{Aut}^*(\mathfrak{n})$  which may arise from all possible nilpotent Lie algebras  $\mathfrak{n}$ .

**Theorem 8.1** [7]. *Let  $k$  be a field of characteristic 0 and let  $A \subset GL(n, k)$  be an algebraic subgroup,  $n \geq 2$ . Then there exists an  $n$ -generator nilpotent Lie algebra  $\mathfrak{n}$  over  $k$  such that  $\text{Aut}^*(\mathfrak{n}) = A$ .*

### COROLLARY 8.2

*There exists a simply connected nilpotent Lie group  $M$  with a lattice such that  $\text{Aut}^*(M) = 1$ , i.e.  $\phi(m)m^{-1} \in [M, M]$  for all  $m \in M$  and any  $\phi \in \text{Aut}(M)$ .*

In particular, this means that  $\text{Aut}(M)$  is unipotent and  $M$  is a so-called characteristically nilpotent Lie group. It may be noted that any lattice in such a group is

“superrigid” in the sense that it is not contained in any other simply connected solvable Lie group as a lattice. Now we may give an example of a rigid lattice  $\Gamma$  in a solvable Lie group  $G$  of (I)-type. This lattice will be constructed with the help of a weakly rigid lattice in a splittable Lie group of (I)-type from example 2.8 and a “superrigid” lattice from the corollary above.

**Example 8.3.** Let us consider the nilpotent Lie group  $M$  with a lattice  $\Gamma'$  from corollary 8.2. We may assume without loss of generality that there exist a one-parameter subgroup  $H \subset M$  and a normal subgroup  $M' \subset M$ ,  $[M, M] \subset M'$  such that  $M = H \cdot M'$  and  $\Gamma' = (\Gamma' \cap H) \cdot (\Gamma' \cap M')$ . Let the element  $h \in H$  generate the cyclic group  $\Gamma' \cap H$ , i.e.  $\Gamma' \cap H = Z(h)$ .

Now put  $U = \mathbf{R}^2 \times M$ . Then  $\text{Aut}(U)$  contains the circle  $SO(2)$  which acts by orthogonal rotations on  $\mathbf{R}^2$  and trivially on  $M$ . Let  $L$  be a one-parameter subgroup in the cylinder  $SO(2) \times H$  such that  $L \cap H = Z(h^4)$ , and let  $G = L \cdot (\mathbf{R}^2 \times M')$ . It is a straightforward verification that

- (1)  $G$  is a simply connected solvable Lie group of (I)-type
- (2)  $U$  is the unipotent hull of  $G$  and  $G_S = SO(2) \cdot U \subset \text{Hol}(U)$
- (3) The nilradical  $N$  of  $G$  is equal to  $\mathbf{R}^2 \times M'$  and  $G = L \cdot N$
- (4)  $G$  is nonsplittable, since the action of  $L$  on  $N$  has a nontrivial unipotent part.

Note that the element

$$a^* = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in SO(2)$$

keeps invariant an integer lattice  $\mathbf{Z}^2 \subset \mathbf{R}^2$ . Put  $\Gamma = Z(a^* \times h) \cdot (\mathbf{Z}^2 \times (\Gamma' \cap M'))$ . Then  $\Gamma \cap U = \mathbf{Z}^2 \times Z(h^4) \cdot (\Gamma' \cap M')$  is a proper subgroup in  $\mathbf{Z}^2 \times \Gamma'$ . Let us prove that  $\Gamma$  is rigid in  $G$ .

In view of corollary 3.9, it suffices to prove that  $\hat{\phi}(G) = G$  for any canonical map  $\hat{\phi}: \text{Hol}(U) \rightarrow \text{Hol}(U)$  such that  $\hat{\phi}(\Gamma) \subset G$ . So let  $\phi \in \text{Aut}(U)$  and let  $\hat{\phi}(an) = \phi a \phi^{-1} \phi(n)$  for all  $a \in \text{Aut}(U)$ ,  $n \in U$ . If  $\hat{\phi}(\Gamma) \subset G$ , then  $\hat{\phi}(T_\Gamma) = \phi T_\Gamma \phi^{-1} \subset T_G$ . But  $T_\Gamma = Z(a^*)$  is the unique cyclic subgroup of  $T_G = SO(2)$  of order 4, and hence  $\phi$  normalizes  $T_\Gamma$ . Let us prove that  $\phi$  commutes with  $T_\Gamma$ .

In fact, since  $\phi$  normalizes  $T_\Gamma$ , it follows that  $\phi$  keeps invariant the weight spaces  $\mathbf{R}^2$  and  $M$  of the action of  $T_\Gamma = Z(a^*)$  on  $U = \mathbf{R}^2 \times M$ . Assume that  $\hat{\phi}(a^*h) = a^{*-1} \phi(h)$ . Since  $\hat{\phi}(a^*h) \in G$ , it follows that  $\phi(h) = h^k m$  for some  $m \in M'$  and  $k \in \mathbf{Z}$  such that  $k + 1 \in 4\mathbf{Z}$ . But the restriction  $\phi: M \rightarrow M$  defines an automorphism of  $M$  and, hence,  $\phi(h) \in hM'$ . This contradiction proves that  $\phi$  commutes with  $T_\Gamma$ .

But the centralizers of  $T_\Gamma$  and  $T_G$  in  $\text{Aut}(U)$  obviously coincide. Consequently,  $\hat{\phi}(an)(an)^{-1} = \phi(n)n^{-1}a^{-1} \in a(\mathbf{R}^2 \times M')a^{-1} = N \subset G$  for all  $a \in T_G$ ,  $n \in U$ . In particular,  $\hat{\phi}(g)g^{-1} \in G$  for all  $g \in G$ , and hence  $\hat{\phi}(G) = G$ . This proves rigidity of  $\Gamma \subset G$ .

**Remark 1.** The lattice  $\Gamma$  is virtually nilpotent, since it has the nilpotent subgroup

*Remark 2.* It would be interesting to construct an explicit example of a rigid lattice in a solvable Lie group of (*I*)-type. Trying to do this we constructed 6-dimensional nilpotent Lie algebra  $\mathfrak{n}$  with three generators  $x, y, z$  satisfying the relations  $[x, y] = [[x, z], x] = [[y, z], y]$ . However, the group  $\text{Aut}(\mathfrak{n})$  turned out to be larger than required for the desired example. We would like to mention here, that this algebra is not included in the well-known classification of nilpotent 6-algebras (cf. for instance, [18]). Therefore, the classification is incomplete.

## Acknowledgements

The work was partially supported by grants from the American Mathematical Society and the International Science Foundation of G Soros. The author would like to thank the Mathematical Sciences Research Institute at Berkeley and the Tata Institute of Fundamental Research, Bombay, where he prepared this work.

## References

- [1] Auslander L, Bieberbach's theorem on space groups and discrete uniform subgroups of Lie groups. I, *Ann. Math.* **71** (1960) 579–590
- [2] Auslander L, An exposition of the structure of solvmanifolds, *Bull. Am. Math. Soc.* **79** (1973) 227–261
- [3] Auslander L and Brezin J, Almost algebraic Lie algebras, *J. Algebra* **8** (1968) 295–313
- [4] Auslander L and Green L, G-induced flows, *Am. J. Math.* **88** (1966) 43–60
- [5] Auslander L, Green L and Hahn F, Flows on homogeneous spaces (Princeton: Univ. Press.) (1963)
- [6] Auslander L and Tolimieri R, Splitting theorems and the structure of solvmanifolds, *Ann. Math.* (2) **92** (1970) 164–173
- [7] Bryant R and Groves J, Algebraic groups of automorphisms of nilpotent groups and Lie algebras, *J. London Math. Soc.* **33** (1986) 453–466
- [8] Gorbatsevich V V, Vinberg E B and Shvartsman O V, Discrete subgroups of Lie groups, in *Itogi nauki i tekhniki. VINITI* **21** (1988) 5–120
- [9] Grunewald F and Segal D, Reflections on the classification of torsion-free nilpotent groups in *Group theory: essays for Philip Hall* (London: Academic Press) (1984) 121–158
- [10] Hochschild G and Mostow G D, Representations and representative functions of Lie groups, *Ann. Math.* **66** (1957) 495–542
- [11] Mostow G D, Representative functions on discrete groups and solvable arithmetic subgroups, *Am. J. Math.* **92** (1970) 1–32
- [12] Mostow G D, Factor spaces of solvable groups, *Ann. Math.* **60** (1954) 1–27
- [13] Mostow G D, Cohomology of topological groups and solvmanifolds, *Ann. Math.* **73** (1961) 20–48
- [14] Malcev A I, Solvable Lie algebras, *Izv. Akad. Nauk. SSSR Ser. Mat.* **9** (1945) 329–352.
- [15] Malcev A I, On a class of homogeneous spaces, *Izv. Akad. Nauk SSSR. Ser. Mat.* **13** (1949), 9–22
- [16] Milovanov M V, On the extension of automorphisms of uniform discrete subgroups of solvable Lie groups, *Dokl. AN BSSR* **17** (1973) 892–895
- [17] Milovanov M V, Description of solvable Lie groups with a given uniform subgroup, *Mat. Sb.* **113** (1980) 98–117
- [18] Morozov V V, Classification of nilpotent Lie algebras of order 6, *Izv. Vys. Uchebn. Zaved. Matematika* **4** (1958) 161–171
- [19] Platonov V P and Milovanov M V, Determination of algebraic groups by arithmetic subgroups, *Dokl. Akad. Nauk SSSR* **279** (1973) 43–46
- [20] Raghuathan M S, Discrete subgroups of Lie groups, (Berlin, Springer-Verlag) (1972)
- [21] Saito M, Sur certains groupes de Lie resolubles. I, II, *Sci. Pap. Coll. Gen. Ed. Univ. Tokyo* **7** (1957) 1–11; **2**, 157–168
- [22] Wang H C, On the deformations of lattices in a Lie group, *Am. J. Math.* **92** (1970) 389–397



## Some remarks on the Jacobian question

SHREERAM S ABHYANKAR

Mathematics Department, Purdue University, West Lafayette IN 47907, USA

Notes by MARIUS VAN DER PUT and WILLIAM HEINZER

Updated by AVINASH SATHAYE

MS received 16 September 1993; revised 21 March 1994

**Abstract.** This revised version of Abhyankar's old lecture notes contains the original proof of the Galois case of the  $n$ -variable Jacobian problem. They also contain proofs for some cases of the 2-variable Jacobian, including the two characteristic pairs case. In addition, proofs of some of the well-known formulas enunciated by Abhyankar are actually written down. These include the Taylor Resultant Formula and the Semigroup Conductor formula for plane curves. The notes are also meant to provide inspiration for applying the expansion theoretic techniques to the Jacobian problem.

**Keywords.** Plane curves; Jacobian; automorphism; Newton-Puiseux expansion.

### Introduction

What follows may best be described by altering the famous phrase of Zariski and Samuel – *this paper is the unborn parent of its child*. These notes, based on Abhyankar's lectures, were originally prepared by van der Put in 1972. Since he had to rush back to the Netherlands in the middle of their preparation, they were worked on by William Heinzer to some extent. But there were still unfinished parts and as a result, they stayed buried in Abhyankar's private papers for 21 years. Later, Abhyankar's work on the Jacobian problem was partly reported in the Tata Institute Notes by Balwant Singh [A1]. Avinash Sathaye rearranged the old notes and added some additional topics based on further discourses by Abhyankar. Paul Eakin and David Shannon have also contributed to the current form of these notes by their critical proof-reading and suggestions. Special thanks are also given to the referee whose comments have significantly improved the logical clarity of this exposition.

In the old tradition of Abhyankar's lecture-notes, these notes are also unread by Abhyankar in their final form. The responsibility for the exposition, therefore, rests with the note-takers.

Much has been published about the Jacobian problem in the meantime, but except for the Balwant Singh Notes, the novel technique of using a combination of Newton-Puiseux expansions at infinity and studying the resulting value-semigroups did not get much exposure. Our aim in reviving these old notes is to renew interest in these methods, which, to paraphrase Abhyankar's own words, "never really got stuck, but only got very tiring". Perhaps, this time one of the methods will be carried through!

The notes' main values are historical and motivational. We have tried to stay close to the original notes and hence there are no brand new theorems. In fact, they do not even give all the known results as found in [A1]. There are, however, proofs of some theorems enunciated elsewhere without proofs as described below.

These notes have two major parts.

In § 1 and § 2, we discuss the  $n$ -variable Jacobian problem. Thus, given  $n$  polynomials  $(u) = (u_1, \dots, u_n)$  in  $n$  variables  $(x) = (x_1, \dots, x_n)$ , we assume that their jacobian  $J_x(u) = \left( \frac{du_i}{dx_j} \right)$  is a nonzero constant in the ground field  $k$ . The problem is to deduce that  $k[x_1, \dots, x_n] = k[u_1, \dots, u_n]$ .

Of course, if the field has positive characteristic, it is well known that the Jacobian problem as stated has a negative answer. One standard example is

$$u_1 = x_1 + x_1^p, \quad u_2 = x_2$$

where  $p$  is the characteristic. Clearly, similar examples exist in all dimensions. So we assume that  $k$  has characteristic 0.

In general, the problem is still unsolved.

We give a relatively simple proof under the additional assumption that the field extension  $k(x)$  over  $k(u)$  is essentially Galois. Explicitly this means that either the fields are equal or that there is a (nontrivial) Galois extension  $L$  of  $k(u)$  contained in  $k(x)$ . A topological proof of this fact was first published by Campbell [Ca]. Most of the arguments presented below have already appeared in [A1]. However, in [A1], only the two dimensional theorem was deduced.

The second part deals with the two dimensional problem.

In § 3, we restrict to the two variables  $x, y$  and take two polynomials over a field  $k$  (of characteristic 0, of course) satisfying the Jacobi condition that  $J_{x,y}(f, g)$  is a nonzero constant. Temporarily, by "degree", let us mean the total degree with respect to  $(x, y)$ . It is evident that the jacobian of the highest degree parts of  $f, g$  will either be 0 or will give the highest degree terms of the jacobian  $J_{x,y}(f, g)$ . Moreover, in the latter case, the degree of the jacobian is exactly equal to the sum of the degrees of  $f, g$  minus the sum of the degrees of  $x, y$ . Much can be deduced by generalizing the idea of a degree by assigning weight  $a$  to  $x$  and weight  $b$  to  $y$  so that a monomial  $x^i y^j$  has weight  $ai + bj$ . We illustrate the use by disposing of the case when the "usual degrees" of  $f, g$  are either coprime or when their GCD is prime.

The last two sections were not part of the original notes.

In § 4, we discuss yet another viewpoint, also developed in [A1]. For this, it is convenient, though not quite necessary, to arrange that  $f, g$  are monic in, say  $y$ . By the basic "two points at infinity" Lemma (3.5), we get to assume that either the proof is finished or we may assume the usual degree forms to be powers of  $y^s(y + cx)^t$  for some  $s, t$ . Then, we may think of the pair  $(f, g)$  as giving a parametrized plane curve over the ground field  $k(x)$ , where  $y$  is thought of as the parameter of the curve. This fits the mold of the Epimorphism Theorem calculations. One of the simplest observations deducible from this is that the Jacobian problem is equivalent to proving that this curve is nonsingular. We begin by giving the Taylor Resultant formula of Abhyankar, developed for this purpose in 1972, which calculates the "conductor" of this plane curve directly as a polynomial in  $x, f, g$ . This formula has since been stated without proof in [A2, page 153] and we take this opportunity to write down the

proof, in view of the interest in the formula. In the remaining part of §4, we write another formula for the conductor in terms of the special generators of value-semigroups developed by using the "Expansion Techniques" as in [A1]. This supplies most of the details for the formula in [A2, page 169].

In §5, we expand on the theme of §3 and give a solution of the Jacobian problem for the case of "two characteristic pairs". The results generalize the main results from §3 and further illustrate the expansion techniques as applied to the problem.

Abhyankar has unpublished results disposing of cases when the plane curve  $(f, g)$  has "a small number" of singularities, but the calculations are too messy to be included in a paper of this nature.

## 1. The Jacobi condition

*Notation.* Let  $k$  be a field and let  $x_1, \dots, x_n$  be  $n$  indeterminates over  $k$ . Given polynomials  $u_1, \dots, u_n \in k[x_1, \dots, x_n]$ , we consider the Jacobian of  $u_1, \dots, u_n$  with respect to  $x_1, \dots, x_n$ :

$$J_x(u) = \det \left( \frac{\partial u_i}{\partial x_j} \right).$$

We say that the polynomials  $(u)$  satisfy the JACOBI-CONDITION, or briefly JC, if we have that  $J_x(u) = \theta$ . Here  $\theta$  stands for "Abhyankar's Nonzero", namely, any suitable nonzero element of the ground field  $k$ . Note that  $\theta$  may denote different numbers depending on the context, perhaps, even in the same equation.

Sometimes, we may replace the field  $k$  by a suitable domain and we may need the **generalized JC** which states that  $J_x(u)$  is a unit in  $k$ .

Let  $\Omega$  denote the universal module of differentials of the polynomial ring  $k[x_1, \dots, x_n]$  over  $k$ . Recall that  $\Omega$  is a vector space generated by  $dx_1, \dots, dx_n$  over  $k$ . By  $\wedge^n \Omega$ , as usual, we will denote the  $n$ th exterior power of  $\Omega$ , which is a one-dimensional vector space generated by  $dx_1 \wedge \dots \wedge dx_n$ . Thus, another way of describing the Jacobian is by writing

$$du_1 \wedge \dots \wedge du_n = J_x(u) dx_1 \wedge \dots \wedge dx_n \in \wedge^n \Omega.$$

We can conveniently abbreviate this as  $du = J_x(u) dx$ .

We need some standard simple facts about the universal module of differentials to reformulate JC<sup>1</sup>.

### Properties of Differentials

(1) Let  $A$  be a ring and  $B$  an  $A$ -algebra. Then there exists a  $B$ -module  $\Omega_{B/A}$  and an  $A$ -derivation  $d: B \rightarrow \Omega_{B/A}$  such that for every  $A$ -derivation  $D$  of  $B$  into a  $B$ -module  $M$ ,

<sup>1</sup> This material was part of the original notes and is left intact for historical reasons. For readers familiar with these concepts, it suffices to note that: for an  $A$ -algebra  $B$  let  $\Omega_{B/A}$  denote, as usual, the universal

there is an unique  $B$ -linear map  $\alpha: \Omega_{B/A} \rightarrow M$ , such that  $D = \alpha \circ d$ . Indeed, this is the defining property of the universal module  $\Omega_{B/A}$ .

(2) If  $B = A[x_1, \dots, x_n]$ , then  $\Omega_{B/A} = Bdx_1 + \dots + Bdx_n$ , where  $\{dx_1, \dots, dx_n\}$  is a free basis and  $d$  is given by  $df = \sum_i \frac{\partial f}{\partial x_i} dx_i$ .

(3) If  $S \supset T$  are respectively multiplicative sets in  $B$  and  $A$ , then

$$\Omega_{S^{-1}B/T^{-1}A} = \Omega_{S^{-1}B/A} = S^{-1}\Omega_{B/A}.$$

(4) If  $I$  is an ideal in  $B$ , then

$$\Omega_{(B/I)/A} = \frac{\Omega_{B/A}}{(I\Omega_{B/A} + BdI)}.$$

(5) If  $B$  is essentially of finite type over  $A$ , i.e.  $B$  is a localization of a finitely generated ring over  $A$ , then  $\Omega_{B/A}$  is a finite  $B$ -module.

(6) For a finitely generated field extension  $L$  of a field  $K$ , the condition that  $\Omega_{L/K} = 0$  is equivalent to  $L$  being a finite separable algebraic extension of  $K$ .

*Proof.* We will briefly indicate the idea of the proof behind these. The first five properties are deduced by formal algebraic manipulations by proving the existence of  $\Omega_{B/A}$  constructively.

For the sixth property above, let us indicate more details. Write, using standard field theory,  $K \subset K_1 \subset K_2 \subset L$ , where  $K_1$  is a pure transcendental extension generated by a transcendence basis of  $L$  over  $K$ ,  $K_2$  is separable algebraic over  $K_1$  and  $L$  is purely inseparable over  $K_2$ .

Suppose, that  $\Omega_{L/K} = 0$ , or, in other words, every  $K$ -derivation of  $L$  into itself, is trivial.

If  $L \neq K_2$ , then there exists a nontrivial  $K_2$ -derivation of  $L$  into  $L$  and this contradicts the hypothesis. So,  $L = K_2$ , and thus  $L$  is separable algebraic over  $K_1$ .

If  $K \neq K_1$ , then, set one of the transcendence basis to be  $x$ . The  $K$ -derivation  $\partial/\partial x$  of  $K_1$  extends to the separable algebraic extension field  $L$  and again we get a nontrivial  $K$ -derivation of  $L$ , a contradiction. Thus  $K = K_1$  and hence  $L$  is separable algebraic over  $K$ .

The converse is obvious!

**Lemma 1.1.** Let  $B = k[x_1, \dots, x_n]$  be a polynomial ring over a field  $k$  and let  $u_1, \dots, u_n$  be polynomials. Set  $A = k[u_1, \dots, u_n]$ . Let  $K = Qt(A)$  and  $L = Qt(B)$  be the respective quotient fields of  $A, B$ .

- (1)  $A = B$  implies  $J_x(u) = \emptyset$ , i.e.,  $0 \neq J_x(u) = c$  for some  $c$  in  $k$ .
- (2)  $L/K$  is a finite separable extension if and only if  $J_x(u) \neq 0$ .
- (3)  $J_x(u) \neq 0$  implies that  $u_1, \dots, u_n$  are algebraically independent over  $k$  and converse holds if  $k$  has zero characteristic.

*Proof.* (1) Since  $A = B$ , both the  $n$ -differentials  $du = du_1 \wedge \dots \wedge du_n$  and  $dx = dx_1 \wedge \dots \wedge dx_n$  are generators of the one-dimensional vector space  $\wedge^n \Omega_{B/k} = \wedge^n \Omega_{A/k}$ . Hence, they differ from each other by a nonzero constant in  $k$ , i.e.  $du = cdx$  where  $0 \neq c \in k$ . Clearly,  $J_x(u) = c$  as stated.

(2) Since  $L$  is a finitely generated extension of  $K$ , we know that  $L/K$  is finite separable if and only if  $\Omega_{L/K} = 0$  if and only if  $\Omega_{L/K} = \Omega_{K/k}$ . The last condition is evidently equivalent to  $u_1, \dots, u_n$  forming another free  $L$ -basis of  $\Omega_{L/k}$  or equivalently,  $du_1, \dots, du_n$ , are independent over  $L$ .

(3) Follows from (2).

**Definition.** Let  $R \subset S$  be local rings with the maximal ideals  $m(R) \subset m(S)$ . We say that  $S$  is **unramified over  $R$**  (or  $S/R$  is **unramified**) if

- (1)  $S/m(S)$  is finite separable over  $R/m(R)$  and
- (2)  $m(R)S + m(S)^2 = m(S)$ .

Note that the second condition implies that  $m(R)S = m(S)$  in case  $m(S)$  is finitely generated.

More generally, given an  $A$ -algebra  $B$  and a prime ideal  $p$  in  $B$ , we say that  $B$  is **locally unramified over  $A$**  at  $p$ , if  $B_p/A_{p \cap A}$  is unramified. We say that  $B$  is **unramified over  $A$** , if  $B$  is locally unramified over  $A$  at all primes  $p$ .

**Lemma 1.2.** Let  $R \subset S$  be local rings with  $m(R) \subset m(S)$ . Assume that  $S$  is essentially of finite type over  $R$ . Then  $S/R$  is unramified if and only if  $\Omega_{S/R} = 0$ .

*Proof.* Suppose that  $S/R$  is unramified. We know that  $\Omega_{S/R}$  is a finite  $S$ -module. Consider the  $R$ -derivation

$$D: S \xrightarrow{d} \Omega_{S/R} \rightarrow \frac{\Omega_{S/R}}{m(S)\Omega_{S/R}} = M, \text{ say.}$$

Now any  $x \in m(S)$  can be written as  $x = \sum_i x_i m_i + \sum_j y_j z_j$ , where  $x_i \in S$ ,  $y_i, z_j \in m(S)$  and  $m_i \in m(R)$ . It is easy to check that  $D(x) = 0$ . Thus  $D$  induces a  $R/m(R)$ -derivation  $\bar{D}: S/m(S) \rightarrow M$ . Since  $S/m(S)$  is finite and separable over  $R/m(R)$  by hypothesis, we get that  $\bar{D} = 0$ . It follows that  $D = 0$  and  $M = 0$ . Thus, by Nakayama's lemma, we get  $\Omega_{S/R} = 0$ .

Now assume that  $\Omega_{S/R} = 0$ . Setting  $S/m(S) = L$  and  $R/m(R) = K$  we can deduce that  $\Omega_{L/K} = 0$ . Thus  $L/K$  is a finite and separable extension. Now the complete local ring  $T = S/(m(R)S + m(S)^2)$  has a coefficient field containing  $K$ . If  $T \neq K$ , then  $T$  has a residue ring of type  $K[x]/(x^2)$ . This last ring has a nontrivial  $K$ -derivation (and hence an  $A$ -derivation)  $x(\partial/\partial x)$ . This contradicts  $\Omega_{S/R} = 0$ , so  $T = K$  or equivalently  $m(S) = m(R)S + m(S)^2$ .

**COROLLARY 1.3.** Let  $B$  be an  $A$ -algebra of essentially finite type. Then  $\Omega_{B/A} = 0$  if and only if  $B$  is locally unramified at every prime ideal  $p$ .

More generally,  $\Omega_{B/A} = 0$  if and only if  $B$  is locally unramified at every maximal ideal  $p$ .

*Proof.* Since  $\Omega_{B/A}$  is finitely generated, we have that  $\Omega_{B/A} = 0$  if and only if  $(\Omega_{B/A})_p = 0$  for all primes  $p$  in  $B$  if and only if  $(\Omega_{B/A})_p = 0$  for all maximal ideals  $p$  in  $B$ . Now we use  $(\Omega_{B/A})_p = \Omega_{B_p/A_{p \cap A}}$  together with the above lemma.

**PROPOSITION 1.4**

Let  $A = k[u_1, \dots, u_n]$  with polynomials  $u_1, \dots, u_n$  in  $B = k[x_1, \dots, x_n]$ , a polynomial ring in  $n$ -variables over a field  $k$ . Then the following are equivalent:

- (1) The polynomials  $u_1, \dots, u_n$  satisfy the Jacobi-condition  $J_x(u) = \emptyset$ .
- (2)  $\Omega_{B/A} = 0$ .
- (3)  $B$  is locally unramified over  $A$  at every prime ideal  $p$  of  $B$ .
- (4)  $B$  is locally unramified over  $A$  at every maximal ideal  $p$  of  $B$ .

*Proof.* Obvious from the above discussion.

**COROLLARY 1.5.** *Let  $A, B$  be as in (1.4) and assume that the Jacobi-condition is satisfied for the polynomials  $u_1, \dots, u_n$ . Let  $\bar{A}$  denote the integral closure of  $A$  in  $B$ . Then we have the following:*

- (1) *For every prime  $p$  in  $B$ ,  $ht(p) = ht(q)$ , where  $q = p \cap A$  and  $B$  is locally unramified over  $A$  at  $p$ . Moreover,  $\hat{B}_p$  is a finite free  $\hat{A}_q$ -module.<sup>2</sup>*
- (2) *If  $p$  is a height 1 prime of  $B$ , then  $\bar{A}_{p \cap \bar{A}} = \bar{B}_p$ .*
- (3) *If  $k$  is algebraically closed and  $m$  is a maximal ideal of  $B$ , then  $B_m = \bar{A}_{m \cap \bar{A}}$ .*
- (4) *Let  $V$  be any dvr (rank 1 discrete valuation ring) of  $Qt(B)$  which contains  $B$ . Then  $V$  is unramified over  $W = V \cap Qt(A)$ .*

*Proof.* (1) We already know that  $B_p$  is unramified over  $A_q$ . This implies that  $\hat{B}_p$  is a finite  $\hat{A}_q$ -module. This, in turn, implies that  $q$  and  $p$  have the same heights. Moreover, both  $\hat{B}_p$  and  $\hat{A}_q$  are regular local rings, so by [Na, (25.16)] we get that  $\hat{B}_p$  is a free  $\hat{A}_q$ -module.

(2)  $\bar{A}_{p \cap \bar{A}}$  is a dvr contained in  $B_p$ , so by maximality of a dvr, it coincides with  $B_p$ .

(3) Using (1), we see that  $A_{m \cap A} = B_m$ . Let  $C = \bar{A}_{m \cap \bar{A}}$ . Then  $C$  is normal, hence analytically normal and hence contained in  $B_m$ . Since both  $C$  and  $B_m$  have the same quotient field, they are equal.

(4) Consider  $R = W[x_1, \dots, x_n] \subset V$  and set  $p = R \cap m(V)$ , where  $m(V)$  is the maximal ideal of  $V$ . Since  $\Omega_{B/A} = 0$ , we get that  $\Omega_{R/W} = 0$  and  $\Omega_{R_p/W} = 0$ . By Lemma 1.2,  $R_p$  is unramified over  $W$ . In particular  $R_p$  is a dvr and hence  $V = R_p$ .

## 2. The Galois case

In this section, we will use the following hypothesis, unless otherwise declared.

**Hypothesis.** Let  $k$  be a field,  $B = k[x_1, \dots, x_n]$  the polynomial ring in  $n$  variables over  $k$  and let  $u_1, \dots, u_n$  be polynomials in  $B$ . Set  $A = k[u_1, \dots, u_n]$ . As before, we say that  $u = (u_1, \dots, u_n)$  satisfies JC (the Jacobi-condition), if  $J_x(u) = \emptyset$ .

Some of our results can be stated and proved under the following more general hypothesis.

**Generalized Hypothesis.** Let  $k$  be normal domain. Assume that  $k$  is prefactorial, i.e. assume that every height 1 prime of  $k$  is the radical of a principal ideal.

Let  $B = k[x_1, \dots, x_n]$  the polynomial ring in  $n$  variables over  $k$  and let  $u_1, \dots, u_n$  be polynomials in  $B$ . Set  $A = k[u_1, \dots, u_n]$ .

<sup>2</sup>Here  $ht$  denotes the usual height of a prime ideal and  $\hat{\phantom{x}}$  denotes the usual completion.

**PROPOSITION 2.1** (Birational Case).

Suppose that  $Qt(A) = Qt(B)$  and  $u$  satisfies JC. Then  $A = B$ , i.e., the Jacobian Theorem holds.

*Proof.* Let  $q$  be any height 1 prime of  $A$ . Then  $q = aA$  for a nonunit  $a \in A$ . Now,  $a$  is a nonunit in  $B$ . Write a factorization  $a = p_1 \cdots p_r$ , where  $p_1 \cdots p_r$  are irreducible in  $B$ . Any one of them, say  $p_1$  generates a height 1 prime ideal  $p = p_1 B$  and  $p \cap A \supset q$ . By Corollary 1.5, we get that  $p \cap A$  has height 1 and hence  $p \cap A = q$ . Hence,  $A_q \subset B_p$  and since both are dvr with the same quotient field, they are equal. Consequently,  $A_q \supset B$ .

Now we have

$$A = \cap \{A_q | q \text{ is a height one prime}\} \supset B \supset A.$$

**COROLLARY 2.2.**

Let  $R \subset S$  be noetherian domains such that:

- (1)  $\Omega_{S/R} = 0$ .
- (2)  $Qt(R) = Qt(S)$  and  $S/R$  is essentially of finite type.
- (3) Any nonunit in  $R$  is a nonunit in  $S$ .
- (4)  $R$  is normal and prefactorial.

Then  $R = S$ .

*Proof.* The only place where the JC was used in the proof of the Proposition 2.1, was in the application of Corollary 1.5 to deduce that the contraction of a height 1 prime has height 1. This can be alternatively proved by the first two conditions of our Corollary; for details see the proof of part 1 of Corollary 1.5.

*Remarks.* (1) Proposition 2.1 remains valid under the generalized hypothesis, provided we assume that  $u$  satisfies the generalized JC also. This follows from Corollary 2.2 after noting that since  $k$  is prefactorial,  $A = k[u_1, \dots, u_n]$  is prefactorial. Such an example where  $k$  is the ring of integers was already discussed by O. Keller [Ke].

(2) The condition that  $R$  be prefactorial is essential. Indeed, take

$$R = k[x, xy, y(1 + xy)] \subset S = k[x, y].$$

Note that  $R$  is isomorphic to  $k[u, v, w]/(uw - v(v + 1))$  and hence  $R$  is easily seen to be regular. To see that  $\Omega_{S/R} = 0$ , note that  $dx, d(xy)$  and  $d(y(1 + xy))$  generate  $\Omega_{S/k}$ . Clearly  $Qt(R) = Qt(S)$  and obviously,  $R \neq S$ . Indeed,  $R$  is not prefactorial. To see this, consider the ideal  $I = (u, 1 + v)R$ . Suppose that it is the radical of a principal ideal in  $R$ . In  $S$  it extends to the ideal  $(x, 1 + xy)S = (1)S$ . Since only units in  $S$  are constants, radical of  $I$  and hence  $I$  itself would be the unit ideal in  $R$ . But the residue class ring  $R/I$  is isomorphic to  $k[w]$ , a contradiction!

**PROPOSITION 2.3** (Galois Case)

Suppose that  $u$  satisfies JC and either  $Qt(A) = Qt(B)$  or that there is a nontrivial tame Galois extension  $L$  of  $Qt(A)$  contained in  $Qt(B)$ , then  $A = B$ .

*Quick Proof.* By Proposition 2.1, we may assume that  $\text{Qt}(A) \neq \text{Qt}(B)$ . Using the fact that the affine  $n$ -space is simply connected (i.e. there are no unramified proper tame extensions of  $k(u_1, \dots, u_n)$ ) and the purity of the branch locus, we deduce that there exists a height 1 prime  $q$  in  $A$  and a dvr  $V$  with quotient field  $L$ , such that  $m(V) \cap A = q$  (where  $m(V)$  is the maximal ideal of  $V$ ) and  $V$  is ramified over  $A_q$ . Since  $L$  is a Galois extension of  $\text{Qt}(A)$ , all extensions of  $A_q$  to  $L$  are ramified over  $A_q$ . Hence, every extension of  $A_q$  to  $\text{Qt}(B)$  is also ramified over  $A_q$ . Now, since  $A$  and  $B$  have the same units (nonzero elements of  $k$ ), we must have at least one prime ideal  $p$  in  $B$  such that  $p \supset q$ . Choosing a minimal prime  $p$  with this property, we get that  $\text{ht}(p) = 1$ . Now by Corollary 1.5  $p \cap A = q$  and  $B_p$  is unramified over  $A_q$ . Since  $B_p$  is evidently an extension of  $A_q$  to  $\text{Qt}(B)$ , we get a contradiction!

## COROLLARY 2.4

*Proposition 2.3 remains valid under the generalized hypothesis, provided we use the generalized JC also.*

*Proof.* Set  $\text{Qt}(k) = K$  and let  $B_1 = K[x_1, \dots, x_n]$  and  $A_1 = K[u_1, \dots, u_n]$ . Then by Proposition 2.3, we get that  $A_1 = B_1$  and hence  $\text{Qt}(B) = \text{Qt}(A)$ . Now, we are done by Corollary 2.2.

*Simpler Proof.* We needed some celebrated theorems above to deduce that unless  $\text{Qt}(B) = \text{Qt}(A)$ , we must have a height 1 prime  $p$  in  $B$ , such that  $B_p$  is ramified over  $A_{p \cap A}$ . We give a simpler proof of this by reducing the proof to some rather well known theory of functions of one variable.

*Bertini Lemma 2.5.* Let  $F$  be an infinite field and let  $K$  be the quotient field of  $F[z_1, \dots, z_n]$ , where  $z_1, \dots, z_n$  are indeterminates over  $F$ . Assume that  $n \geq 2$ . Let  $E$  be a finite separable algebraic extension of  $K$ , such that  $F$  is algebraically closed in  $E$ . Then for "almost all" linear combinations  $y = \lambda_1 z_1 + \dots + \lambda_n z_n$ , we get that  $F(y)$  is algebraically closed in  $E$  and  $E$  is a separable extension of  $F(y)$ .

*Proof.* The separability of  $E$  over  $F(y)$  is well known and we only demonstrate the algebraic closedness. For  $\lambda \in F$  consider fields  $K_\lambda = \overline{F(y_\lambda)}(z_2, \dots, z_n)$ , where the bar — denotes algebraic closure in  $E$  and  $y_\lambda = z_1 + \lambda z_2$ .

Now we claim that if  $K_\lambda = K_\mu$  for some  $\lambda \neq \mu$ , then  $F(y_\lambda)$  is algebraically closed in  $E$ . Assuming the claim for a moment, we note that, since  $F$  is infinite and since there are only finitely many fields in between  $K$  and  $E$ , we have the result for almost all combinations  $z_1 + \lambda z_2$ . The result can then be deduced with a suitable technical interpretation of "almost all".

To prove the claim, set  $w_1 = z_1 + \lambda z_2$ ,  $w_2 = z_1 + \mu z_2$  and  $w_i = z_i$  for  $i \geq 3$ . By assumption

$$\overline{F(w_1)}(w_2, w_3, \dots, w_n) = \overline{F(w_2)}(w_1, w_3, \dots, w_n).$$

Denote this field by  $L$ . Since  $F$  is algebraically closed in  $\overline{F(w_2)}$ , we get that  $\overline{F(w_1)}(w_2, w_3, \dots, w_n)$  is algebraically closed in  $\overline{F(w_2)}(w_1, w_3, \dots, w_n) = L$ . In particular,  $\overline{F(w_1)}$  is algebraically closed in the field  $L$ . It follows that  $F(w_1) = \overline{F(w_1)}$  as claimed.



## COROLLARY 2.6

With the notation and assumptions of the Bertini Lemma (2.5), there exist  $y_1, \dots, y_n$ , linear combinations of  $z_1, \dots, z_n$ , such that  $F[y_1, \dots, y_n] = F[z_1, \dots, z_n]$  and  $F(y_2, \dots, y_n)$  is algebraically closed in  $E$ .

*Proof.* Apply (2.5) a number of times.

**Lemma 2.7.** Let  $K$  denote the quotient field of  $F[y]$ . Let  $E$  be a finite separable algebraic extension of  $K$  such that  $F$  is algebraically closed in  $E$ . Suppose that  $E/K$  is tamely ramified. If no height one prime of  $F[y]$  is ramified in  $E$  then  $E = K$ .

*Proof.* The algebraic closure  $\bar{F}$  of  $F$  and  $E$  are linearly disjoint over  $F$ . Hence after replacing  $F$  by  $\bar{F}$ ,  $K$  by  $\bar{F}(y)$  and  $E$  by  $E(\bar{F})$  we are reduced to the case where  $F$  is algebraically closed. The canonical divisor for both  $E$  and  $K$  is given by  $dy$  and according to [Ch, p. 106, Corollary 2] we have the formula

$$\deg_E(dy) = \deg_K(dy)[E:K] + \deg(\mathcal{D}_{E/K})$$

where  $\mathcal{D}_{E/K}$  denotes the different of  $E/K$ . The only ramified discrete valuation of  $K/F$  is  $F[y^{-1}]_{(y-1)}$ . Since  $E/K$  is tamely ramified it follows that

$$\deg(\mathcal{D}_{E/K}) \leq [E:K] - 1.$$

Let  $g$  be the genus of  $E/F$ . Then the above formula yields

$$-2 \leq 2g - 2 = -2[E:K] + \deg(\mathcal{D}_{E/K}) \leq -[E:K] - 1.$$

Hence  $[E:K] = 1$  and  $E = K$ .

## PROPOSITION 2.8

Let  $K$  denote the quotient field of the polynomial ring  $k[z_1, \dots, z_n]$  and  $E$  a finite separable algebraic extension of  $K$  such that  $E/K$  is tamely ramified and  $k$  is algebraically closed in  $E$ . If  $E \neq K$ , then some height one prime of  $k[z_1, \dots, z_n]$  ramifies in  $E$ .

*Proof.* Using (2.6) we consider  $F[y]$  where  $F = k(y_2, \dots, y_n)$  and  $y = y_1$ . According to (2.7) some height one prime  $p$  of  $F[y]$  ramifies in  $E$ . Then also  $q = p \cap k[y_1, \dots, y_n] = p \cap k[z_1, \dots, z_n]$  has height one and ramifies in  $E$ .

*Remark.* (2.8) can replace the use of "purity of branch locus" and "simple connectedness" in the proof of (2.3).

## 3. Some results in two variables

*Notation.* In this section we suppose that  $k$  is an algebraically closed field of characteristic 0 and we study polynomials  $f, g \in k[x, y]$  satisfying the Jacobi-condition JC:

$$df \wedge dg = \mp dx \wedge dy$$

where, as already explained,  $\ominus$  is the “Abhyankar’s Nonzero”, i.e. some nonzero element of  $k$ .

For  $f = \sum f_{i,j} x^i y^j$  we consider  $\text{supp}(f) = \{(n, m) \in \mathbb{R}^2 \mid f_{n,m} \neq 0\}$ .

The boundary polygon of the smallest convex set in  $\mathbb{R}^2$  containing  $\text{supp}(f)$  will be called  $N(f)$  = the Newton polygon of  $f$ . (This concept is somewhat similar to ordinary Newton polygon, but certainly not the same.) Let  $E(f)$  denote the set of vertices of  $N(f)$ , or equivalently,  $E(f)$  is the set of extreme points of  $N(f)$ .

In order to show that the Jacobi-condition for  $f$  and  $g$  implies a similarity of  $N(f)$  and  $N(g)$  we introduce degree functions and gradings of  $k[x, y]$ .

Given coprime integers  $a, b \in \mathbb{Z}$  (i.e. integers satisfying  $a\mathbb{Z} + b\mathbb{Z} = \mathbb{Z}$ ), we form a grading  $k[x, y] = \sum_{n=-\infty}^{\infty} H_{a,b}^n$  where  $H_{a,b}^n$  is the  $k$ -vector space generated by all monomials  $x^i y^j$  such that  $ai + bj = n$ . Every nonzero  $f \in k[x, y]$  is written as  $\sum f_i$ , where  $f_i \in H_{a,b}^i$ . The corresponding  $(a, b)$ -degree  $\Delta_{a,b}$  is defined by  $\Delta_{a,b}(f) = \max\{i \mid f_i \neq 0\}$ . The degree form of  $f$  with respect to  $(a, b)$  is by definition equal to  $f_n$  with  $n = \Delta_{a,b}(f)$ . We use the notation  $f_{a,b}^+$  to denote it. If  $f = 0$  then usually its *degree* is taken to be undefined or sometimes equal to any desired number. The degree form is similarly, undefined or 0.

The reference to the weight  $a, b$  may be dropped, if the weights are clear from the context.

It is not hard to see that  $(n, m) \in E(f)$  is equivalent to the existence of an  $(a, b)$ -degree such that  $f_{a,b}^+ = \ominus x^n y^m$ . Moreover, every side of  $N(f)$  corresponds to an  $(a, b)$ -degree for which  $f_{a,b}^+$  is not a constant multiple of a monomial.

On  $\wedge^2 \Omega_{k[x,y]/k}$  we introduce a similar grading associated with  $(a, b)$  by means of

$$\wedge^2 \Omega_{k[x,y]/k} = \sum H_{a,b}^n dx \wedge dy.$$

The corresponding degree-function is again denoted by  $\Delta_{a,b}$ . Further, for any  $w \in \wedge^2 \Omega_{k[x,y]/k}$  we let  $w_i$  denote its homogeneous part (with respect to  $(a, b)$ ) of order  $i$ .

The definitions concerning weights can easily be generalized to arbitrary real numbers  $(a, b)$ . For analyzing Newton diagrams, however, we only need integer weights or sometimes, for clarity, we may use rational weights.

*Lemma 3.1. Let  $a, b$  be coprime integers. We write  $H^n$  for  $H_{a,b}^n$  and  $\Delta$  for  $\Delta_{a,b}$ . We will also drop reference to  $a, b$  from the various degree-form notations. Then we have:*

- (1)  $dH^n \wedge dH^m \subset H^{n+m-a-b} dx \wedge dy$ .
- (2) For any  $f, g \in k[x, y]$  one has

$$(df \wedge dg)_{\Delta(f) + \Delta(g) - \Delta(xy)} = df^+ \wedge dg^+.$$

*In particular,  $\Delta(f) + \Delta(g) - \Delta(xy) \geq \Delta(df \wedge dg)$ . Strict inequality holds if and only if  $df^+ \wedge dg^+ = 0$ .*

- (3) If  $f, g \in k[x, y]$  are nonzero polynomials, homogeneous with respect to  $(a, b)$ , then  $df \wedge dg = 0$  implies  $f^{\Delta(g)} = \ominus g^{\Delta(f)}$ .
- (4) If  $f, g \in k[x, y]$  satisfy  $d(xf) \wedge d(g) = \ominus dx \wedge dy$ , then  $0 \neq f \in k$  and  $g = \ominus y + p(x)$  for some polynomial  $p$ . In particular,  $k[f, g] = k[x, y]$ .
- (5) If  $f, g \in k[x, y]$  satisfy  $df \wedge dg = \ominus dx \wedge dy$  and  $f$  is homogeneous with respect to  $(a, b)$ , then  $k[f, g] = k[x, y]$  and either  $\deg(f) = 1$  or  $f = \ominus x + p(y)$  or  $f = \ominus y + p(x)$  for some polynomial  $p$ .

*Proof.*

(1) For monomials  $x^{i_1}y^{j_1}$  and  $x^{i_2}y^{j_2}$  we have

$$d(x^{i_1}y^{j_1}) \wedge d(x^{i_2}y^{j_2}) = (i_1j_2 - i_2j_1)x^{i_1+i_2-1}y^{j_1+j_2-1}dx \wedge dy.$$

The statement in (1) follows easily from that.

(2) Let  $n = \Delta(f)$  and  $m = \Delta(g)$ . Write

$$f = \sum_{i \leq n} f_i \text{ and } g = \sum_{j \leq m} g_j.$$

Clearly

$$df \wedge dg = \sum_d \sum_{i+j=d} df_i \wedge dg_j$$

and the term  $\sum_{i+j=d} df_i \wedge dg_j$  is homogeneous (with respect to  $(a, b)$ ) of order  $d - a - b = d - \Delta(xy)$ . This makes statement (2) obvious.

(3) Consider the 1-form  $w = -(by)dx + (ax)dy$ . It has the property that for homogeneous  $h_1, h_2 \in k[x, y]$  (homogeneous with respect to  $(a, b)$ ) and  $h_2 \neq 0$  the following formula holds:

$$d\left(\frac{h_1}{h_2}\right) \wedge w = (\Delta(h_1) - \Delta(h_2))\left(\frac{h_1}{h_2}\right)dx \wedge dy.$$

Apply this to  $v = f^{\Delta(g)}g^{-\Delta(f)}$ . Then it gives  $dv \wedge w = 0$ .

We want to show that  $v \in k$ . If not, then  $dv \neq 0$  and is linearly dependent with  $w$  (over the field  $k(x, y)$ ). So  $dv = hw$  with  $0 \neq h \in k(x, y)$ . Also  $dv \wedge df = dv \wedge dg = 0$ , hence  $df \wedge w = dg \wedge w = 0$ . This implies  $\Delta(f) = \Delta(g) = 0$  and  $v = 1$ . Contradiction! Hence  $v \in k$  and we have  $f^{\Delta(g)} = \Theta g^{\Delta(f)}$ . Since  $f$  and  $g$  are actually polynomials,  $\Delta(f)\Delta(g) \geq 0$  and (3) follows. We remark that the statement of (3) becomes uninteresting for  $\Delta(f) = \Delta(g) = 0$ .

(4) Assume first that  $f$  is nonconstant. Without loss of generality we may assume that  $g(0, 0) = 0$ . Then neither  $f$  nor  $g$  is divisible by  $x$  since in either case  $df \wedge dg$  would be divisible by  $x$ . Write  $g = xp(x, y) + q(y)$  where  $q \neq 0$ ,  $q(0) = 0$ . Take  $(a, b)$  such that  $a < 0$  and  $b = 1$ . As usual, by  $a < 0$ , we mean  $a$  is assumed to be sufficiently negative. Thus, for  $a < 0$ , we have  $\Delta(xp(x, y)) < 0$  and  $\Delta(q(y)) = \deg_y q = s > 0$ . Hence  $g_{a,b}^+ = \Theta y^s$ . A similar argument shows that  $f_{a,b}^+ = \Theta y^t$  ( $t > 0$ ) for  $a < 0$  and  $b = 1$ . Hence  $(xf)_{a,b}^+ = \Theta xy^t$ . From (2) it follows that

$$d(xf)_{ab}^+ \wedge d(g_{ab}^+) = vdx \wedge dy \text{ with } v \in k.$$

This is a contradiction. Hence  $f \in k$ ,  $f \neq 0$ . The rest of (4) follows easily.

(5) Write  $g = \sum g_i$ . It follows from (1) that for some  $j$ ,  $df \wedge dg_j = \Theta dx \wedge dy$ . Put  $h = g_j$ . From (2) it follows that  $\Delta(fh) = \Delta(xy)$ , hence  $fh = \sum \lambda_{\alpha\beta} x^\alpha y^\beta$  and  $\text{supp}(fh) \subset \{(\alpha, \beta) | (\alpha - 1)a + (\beta - 1)b = 0\}$ . We suppose (as we may) that  $f(0, 0) = 0$  and  $h(0, 0) = 0$ . If  $x$  or  $y$  divides  $fh$ , then (4) yields  $k[f, h] = k[x, y]$ . It follows that  $g = h + p(f)$  for some polynomial  $p$  and consequently  $k[f, g] = k[x, y]$ .

If neither  $x$  nor  $y$  divides  $fh$ , then there exists  $n > 0$  and  $m > 0$  with  $(n, 0), (0, m) \in \text{supp}(fh)$ . Hence  $(n - 1)a = b$  and  $a = (m - 1)b$ . So either  $a = b = 1$  or  $a = b = -1$ . In both cases  $f$  and  $h$  are linear expressions in  $x$  and  $y$ . Clearly  $k[f, h] = k[x, y]$  and  $k[f, g] = k[x, y]$ . The rest of the statement is easily checked.

*Remark.*

Part (2) above leads to the following natural.

**Definition.** We say that two polynomials  $f, g$  are  $(a, b)$ -related if

$$\Delta_{a,b}(f) + \Delta_{a,b}(g) - \Delta_{a,b}(xy) > \Delta_{a,b}(df \wedge dg)$$

and they are said to be  $(a, b)$ -unrelated otherwise.

We set the  $(a, b)$ -deficiency of  $(f, g)$  to be

$$\delta_{a,b}(f, g) = \Delta_{a,b}(f) + \Delta_{a,b}(g) - \Delta_{a,b}(xy) - \Delta_{a,b}(df \wedge dg).$$

Thus part (3) says that  $(a, b)$ -relatedness is equivalent to the  $(a, b)$ -degree-forms being powers of each other, or equivalently, to the  $(a, b)$ -deficiency being positive.

Part (4) of the above lemma can also be stated as: If  $f, g \in k[x, y]$  satisfy the Jacobi-condition and if there exists a factor  $p$  of  $fg$  and a  $q$  with  $k[p, q] = k[x, y]$ , then  $k[f, g] = k[x, y]$ .

### PROPOSITION 3.2.

(Similarity of Newton polygons) Let  $f, g \in k[x, y]$  satisfy  $df \wedge dg = \ominus dx \wedge dy$  and  $f(0, 0) \neq 0 \neq g(0, 0)$ . Let their total degrees with respect to  $(x, y)$  be respectively  $n, m$ . Assume  $n \geq 2$  and  $m \geq 2$ . Then:

- (1) If  $(a, b)$  is a degree such that  $a \leq 0$  and  $b \leq 0$ , then  $\text{supp}(f_{a,b}^+)$  is either  $\{(0, 0)\}$  or lies entirely on the  $x$ -axis or the  $y$ -axis. The same holds for  $g$ .
- (2) There exists a non-zero constant  $c \in k$  such that for every degree  $(a, b)$  with  $a > 0$  or  $b > 0$  one has

$$(f_{a,b}^+)^m = c(g_{a,b}^+)^n.$$

In particular  $\Delta_{a,b}(f^m - cg^n) < \max(m\Delta_{a,b}(f), n\Delta_{a,b}(g))$  for all degrees  $(a, b)$  with  $a > 0$  or  $b > 0$ .

- (3)  $mN(f) = nN(g)$  and  $mE(f) = nE(g)$ . The Newton polygons of  $f$  and  $g$  look like as in figure 1.

- (4) Let  $(a, b)$  be a degree such that  $a > 0$  or  $b > 0$ . Then there are  $(a, b)$ -homogeneous elements  $h_1, h_2 \in k[x, y]$  such that  $df_{a,b}^+ \wedge dh_1 = h_2 dx \wedge dy$ ,  $\Delta(f_{a,b}^+) + \Delta(h_1) = \Delta(xy) + \Delta(h_2)$  and for some  $p > 0$ ,  $(f_{a,b}^+)^p = \ominus h_2^n$ .

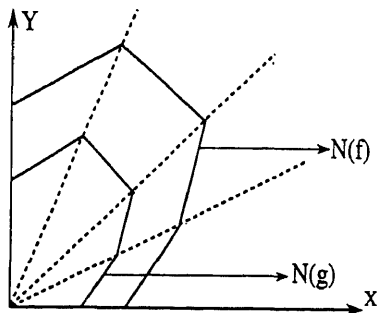


Figure 1.

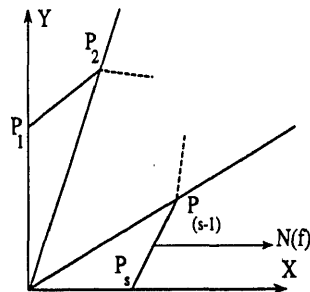


Figure 2.

*Proof.* Our assumptions  $f(0,0) \neq 0$ ,  $\deg f \geq 2$  and  $df \wedge dg = \theta dx \wedge dy$  together with (3.1) part (4) imply that  $f - f(0,0)$  is neither divisible by  $x$  nor  $y$ . So the Newton polygon of  $f$  looks like as in figure 2.

From the diagram, the statements in (1) follow.

Let  $(a, b)$  be a degree with  $a > 0$  or  $b > 0$ . We claim that

$$\Delta_{a,b}(f) + \Delta_{a,b}(g) > \Delta_{a,b}(xy).$$

Consider separately the cases: (i)  $a > 0$ ,  $b \leq 0$ ; (ii)  $a \leq 0$ ,  $b > 0$ ; (iii)  $a \geq b > 0$ ; (iv)  $b \geq a > 0$ .

Case (i):  $\Delta_{a,b}(f) \geq a$ ,  $\Delta_{a,b}(g) \geq a$ , so  $\Delta_{a,b}(fg) \geq 2a > \Delta_{a,b}(xy)$ .

Case (iii): If  $a > b$  the above argument still holds. If  $a = b$ , then  $a = b = 1$  and  $\Delta_{a,b}$  equals the usual total degree on  $k[x, y]$ . It is given that  $\deg(f) + \deg(g) \geq 4 > \deg(xy)$ .

Cases (ii) and (iv) are symmetric to (i) and (iii).

It follows from (3.1) that  $df_{a,b}^+ \wedge dg_{a,b}^+ = 0$  and  $\Delta_{a,b}(f) > 0$ ,  $\Delta_{a,b}(g) > 0$ . Again using (3.1) we get

$$(f_{a,b}^+)^{\Delta_{a,b}(g)} = \theta (g_{a,b}^+)^{\Delta_{a,b}(f)}.$$

In particular, it follows that  $f_{a,b}^+$  is a monomial if and only if  $g_{a,b}^+$  is a monomial. Geometrically, this means that the line through  $(0,0)$  and  $P_i$  (use the numbering of  $E(f)$  as in the figure above) contains a point  $Q_i$  at  $E(g)$  and that all points of  $E(g)$  are obtained in this fashion. Also  $f_{a,b}^+$  is not a monomial (i.e. represents a side of  $N(f)$ ) if and only if  $g_{a,b}^+$  is not a monomial. Hence the sides  $\{P_i, P_{i+1}\}$  and  $\{Q_i, Q_{i+1}\}$  are parallel. It follows that for some rational number  $\lambda \neq 0$  one has  $N(f) = \lambda N(g)$ . Using  $(a, b) = (1, 1)$  one sees that  $\lambda = n/m$ .

Let  $c_i$  be the nonzero constant in  $k$  satisfying  $(f_{a,b}^+)^m = c_i (g_{a,b}^+)^n$  where  $(a, b)$  is chosen such that  $\text{supp}(f_{a,b}^+) = P_i$  and  $\text{supp}(g_{a,b}^+) = Q_i$ . For a degree  $(a, b)$  which corresponds to the side  $(P_i, P_{i+1})$  of the Newton polygon of  $f$  we have  $(f_{a,b}^+)^m = \theta (g_{a,b}^+)^n$ . Comparing the monomials in this equation with maximal or minimal degree in  $x$  or  $y$  one finds  $\theta = c_i = c_{i+1}$ . Hence all  $c_i$  are equal to  $c = c_1$ . This proves (2) and (3).

For (4), put  $T^1(g) = f^m - cg^n$ . We claim that:  $\delta_{a,b}(f, T^1(g)) < \delta_{a,b}(f, g)$ . This can be easily checked from  $df \wedge dT^1(g) = \theta g^{n-1} dx \wedge dy$ . If  $\delta_{a,b}(f, T^1(g)) = 0$ , then  $h_1 = T^1(g)_{a,b}^+$  and  $h_2 = \theta (g_{a,b}^+)^{n-1}$  have the required properties. If  $\delta_{a,b}(f, T^1(g)) > 0$ , then (3.1) yields  $df_{a,b}^+ \wedge dT^1(g)_{a,b}^+ = 0$  and  $(f_{a,b}^+)^p = \theta (T^1(g)_{a,b}^+)^n$  for some  $p > 0$ . Put  $T^2(g) = f^p - \theta T^1(g)^n$ . Then, as before, we deduce that  $\delta_{a,b}(f, T^2(g)) < \delta_{a,b}(f, T^1(g))$ . Since, the deficiency cannot decrease indefinitely, eventually some  $T^r(g)$  has deficiency 0 and we get:

$$df \wedge dT^r(g) = \theta (gT^1(g) \cdots T^{r-1}(g))^{n-1} dx \wedge dy$$

and

$$\Delta(f) + \Delta(T^r(g)) = \Delta(xy) + \Delta(gT^1(g) \cdots T^{r-1}(g))^{n-1}.$$

The elements  $h_1 = T^r(g)_{a,b}^+$  and  $h_2 = \theta ((gT^1(g) \cdots T^{r-1}(g))^{n-1})_{a,b}^+$  have the required properties.

*Remark.* Part (4) of (3.2) can be phrased differently. The condition  $(f_{a,b}^+)^p = \theta h_2^n$  implies that  $f_{a,b}^+$  and  $h_2$  are powers of the same homogeneous element  $C$ . Putting  $D = \theta h_1$ , the equation reads  $dC \wedge dD = C^{t+1} dx \wedge dy$ , with  $t \geq -1$ . We study this equation separately in the sequel to this section.

**Lemma 3.3.** Let  $C$  and  $D$  be  $(a, b)$ -homogeneous elements of  $k[x, y]$ . Write  $C = x^{i_1} y^{j_1} \sigma(z)$  and  $D = x^{i_2} y^{j_2} \tau(z)$  where  $i_1, i_2, j_1, j_2 \in \mathbb{Z}$ ,  $z = x^{-b} y^a$  and  $\sigma, \tau \in k[z]$ . Then

$$dC \wedge dD = CD \left( (i_1 j_2 - i_2 j_1) + z \left( \frac{\Delta(C)\tau'}{\tau} - \frac{\Delta(D)\sigma'}{\sigma} \right) \right) \frac{dx \wedge dy}{xy}.$$

*Proof.* Note that:

$$\left( \frac{dC}{C} \wedge \frac{dD}{D} \right) = \left( \frac{i_1 dx}{x} + \frac{j_1 dy}{y} + \frac{\sigma' dz}{\sigma} \right) \wedge \left( \frac{i_2 dx}{x} + \frac{j_2 dy}{y} + \frac{\tau' dz}{\tau} \right).$$

Using

$$\frac{dx}{x} \wedge \frac{dz}{z} = a \frac{dx \wedge dy}{xy} \text{ and } \frac{dy}{y} \wedge \frac{dz}{z} = b \frac{dx \wedge dy}{xy},$$

it is easy to establish the formula.

**Lemma 3.4.** Let  $C$  and  $D$  be nonzero  $(a, b)$ -homogeneous elements of  $k[x, y]$  such that  $dC \wedge dD = \Theta C^{t+1} dx \wedge dy$  with  $t > 0$ . If  $\Delta(C) \neq 0$  and  $\Delta(C)\Delta(xy) \geq 0$ , then  $C'$  divides  $D$ .

*Proof.* Multiplying by  $C^{t-1}$  the equation becomes  $d(C') \wedge dD = \Theta C^{2t} dx \wedge dy$ . After replacing  $C$  by  $C'$  one sees that it is enough to deal with the case  $t = 1$ . Also, the nonzero constant can be absorbed in  $D$ .

Multiplying both sides of the equation by  $\frac{xy}{CD}$  and using the reductions mentioned above, it is equivalent to

$$\frac{xy}{CD} dC \wedge dD = \frac{Cxy}{D} dx \wedge dy.$$

Using the notation of (3.3) we find that the expression  $\frac{Cxy}{D}$  equals:

$$\frac{x^{i_1+1-i_2} y^{j_1+1-j_2} \sigma}{\tau} = (i_1 j_2 - i_2 j_1) + z \left( \frac{\Delta(C)\tau'}{\tau} - \frac{\Delta(D)\sigma'}{\sigma} \right).$$

Moreover, we can easily prearrange  $\sigma(0) \neq 0$  and  $\tau(0) \neq 0$ , since any factors of  $z$  can be absorbed in the monomials.

The monomial  $x^{i_1+1-i_2} y^{j_1+1-j_2}$ , must then be a rational function of  $z$  and clearly a monomial  $z^r$  for some integer  $r$ . Now clearly, the rational function of  $z$  on the right hand side does not have a pole at  $z = 0$ , so neither does the left hand side. Since  $r$  is the  $z$ -adic order of the left hand side, we get that  $r$  is a nonnegative integer.

Let  $\lambda (\neq 0)$  be any root of  $\sigma$  with multiplicity  $e_1 > 0$  and a root of  $\tau$  with multiplicity  $e_2 \geq 0$ . If we can show  $e_2 \geq e_1$ , then we have proved that  $\sigma$  divides  $\tau$ .

Write  $\sigma = (z - \lambda)^{e_1} \sigma_1$  and  $\tau = (z - \lambda)^{e_2} \tau_1$ . One obtains:

$$\begin{aligned} & z^r (\sigma_1 / \tau_1) (z - \lambda)^{e_1 - e_2} \\ &= (i_1 j_2 - i_2 j_1) + z (\Delta(C)\tau'_1 / \tau_1 - \Delta(D)\sigma'_1 / \sigma_1) + \frac{z(\Delta(C)e_2 - \Delta(D)e_1)}{z - \lambda}. \end{aligned}$$

If  $e_1 > e_2$ , then  $\Delta(C)e_2 = \Delta(D)e_1$ . Using  $\Delta(D) = \Delta(C) + \Delta(xy)$  this yields  $e_1 \Delta(xy) = (e_2 - e_1)\Delta(C)$ , which contradicts the hypothesis  $\Delta(C)\Delta(xy) \geq 0$ .

In order to show that  $C$  divides  $D$  we are left with showing  $i_1 \leq i_2$  and  $j_1 \leq j_2$ . Or, in obvious notation,  $(i_1, j_1) \leq (i_2, j_2)$ .

If  $r = 0$ , then,  $(i_1, j_1) + (1, 1) = (i_2, j_2)$  and we are done.

If  $r > 0$ , then  $(i_1 j_2 - i_2 j_1) = 0$ . Hence for integers  $\lambda_1, \lambda_2$  (not both zero) one has  $\lambda_1(i_1, j_1) = \lambda_2(i_2, j_2)$ . Using again  $\Delta(D) = \Delta(C) + \Delta(xy)$  this yields  $(\lambda_1 - \lambda_2)\Delta(C) = \lambda_2\Delta(xy)$ . If  $\lambda_2\Delta(xy) = 0$  then  $\lambda_1 = \lambda_2$  and we are finished. If  $\lambda_2\Delta(xy) \neq 0$ , then we may take  $\lambda_2 > 0$ . The assumption  $\Delta(xy)\Delta(C) \geq 0$  gives  $\lambda_1 > \lambda_2$  and that implies  $(i_1, j_1) \leq (i_2, j_2)$ .

**Lemma 3.5.** Let  $0 \neq C$  and  $D$  denote  $(a, b)$ -homogeneous elements such that  $dC \wedge dD = Cdx \wedge dy$ . Suppose  $a > 0$  and  $b > 0$ . Then  $C$  has "at most two points at infinity" which means: There are  $(a, b)$ -homogeneous elements  $x_1$  and  $y_1$  with  $k[x, y] = k[x_1, y_1]$  such that  $C = \vartheta x_1^i y_1^j$ .

Moreover,  $D = \vartheta x_1 y_1$  and  $i \neq j$ . The only possibilities are:

- (1)  $a = b = 1$  and  $x_1, y_1$  are linear expressions in  $x$  and  $y$ .
- (2)  $a = 1$  and  $b > 1$  and  $x_1 = x$  and  $y_1 = y + \vartheta x^b$ .
- (3)  $a > 1$  and  $b = 1$  and  $x_1 = x + \vartheta y^a$  and  $y_1 = y$ .
- (4)  $a \geq 1$  and  $b \geq 1$  and  $x_1 = x$  and  $y_1 = y$ .

*Proof.* The formula  $\Delta(D) = \Delta(xy)$  implies that  $D$  must have the form  $\lambda_1 xy + \lambda_2 x^s + \lambda_3 y^t$ .

If  $\lambda_2 \neq 0$  and  $\lambda_3 \neq 0$ , then  $a = b = 1$  and  $s = t = 2$ .  $D$  is obviously equal to  $x_1 y_1$  where  $x_1$  and  $y_1$  are linear in  $x$  and  $y$ . Also,  $D$  cannot be a square, so  $x_1, y_1$  are linearly independent.

If  $\lambda_1 \neq 0$ ,  $\lambda_2 \neq 0$  and  $\lambda_3 = 0$ , then  $D = \vartheta x(y + \vartheta x^{s-1})$ . Since  $y + \vartheta x^{s-1}$  is  $(a, b)$ -homogeneous it follows that  $a = 1$ ,  $b = s - 1$ . We take in this case  $x_1 = x$  and  $y_1 = y + \vartheta x^b$ .

If  $\lambda_1 \neq 0$ ,  $\lambda_2 = 0$ ,  $\lambda_3 \neq 0$ , then similarly  $D = \vartheta x_1 y_1$  with  $x_1 = x + \vartheta y^a$ ,  $y_1 = y$  and  $b = 1$ .

If  $\lambda_1 \neq 0$  and  $\lambda_2 = \lambda_3 = 0$ , then take  $x_1 = x$  and  $y_1 = y$ . In this case  $(a, b)$  can be arbitrary.

Finally,  $\lambda_1 = \lambda_2 = 0$  or  $\lambda_1 = \lambda_3 = 0$  is not possible.

In all cases we have  $D = \vartheta x_1 y_1$ . Write  $C = \sum \lambda_{ij} x_1^i y_1^j$ , with  $\text{supp}(C) \subset \{(i, j) | ai + bj = \Delta(C)\}$ . The equation  $dC \wedge dD = \vartheta C dx_1 \wedge dy_1$  becomes explicit:  $x_1 C_{x_1} - y_1 C_{y_1} = \vartheta C$ . This implies that for some integer  $l$

$$\text{supp}(C) \subset \{(i, j) | i - j = l\}.$$

Hence  $C$  is a monomial in  $x_1$  and  $y_1$ .

**Remarks.** (1) The restriction in (3.5) given by  $a > 0$  and  $b > 0$  is necessary as is shown in the following counterexample:  $(a, b) = (1, 0)$ ,  $C = x^2 y(y + 1)^3$  and  $D = xy(y + 1)$  have the property  $dC \wedge dD = Cdx \wedge dy$  and  $C$  is not a monomial in new variables  $x_1, y_1$ .

(2) Another example:  $(a, b) = (3, -1)$  and  $C = x(xy^3 + 1)^2$  and  $D = xy(xy^3 + 1)$ . Then  $dC \wedge dD = Cdx \wedge dy$ .

(3) **However, the following generalization is valid.** Assume that  $C$  and  $D$  are  $(a, b)$ -homogeneous elements such that  $dC \wedge dD = Cdx \wedge dy$  and the highest  $y$ -degree terms

in  $C$  and  $D$  are unrelated. Assume further that  $a > 0$ . Then the conclusion of the Lemma holds. For a proof, note that the hypothesis implies that the highest  $y$ -degree term in  $D$  must be  $\ominus xy$  since its jacobian with the highest  $y$ -degree term of  $C$  must reproduce itself up to a constant multiplier. It follows that we must have

$$D = \ominus xy + \lambda_2 x^s.$$

The proof is easily finished. Let us further note that in case  $a + b > 0$  and  $b < 0$ , we must have  $\lambda_2 = 0$  and hence  $C, D$  are monomials in  $x, y$  already.

**Theorem 3.6.** *Let  $f, g \in k[x, y]$  satisfy the Jacobi-condition and suppose that their degrees  $n$  and  $m$  satisfy  $n \nmid m$  and  $m \nmid n$ . Then  $\text{GCD}(n, m)$  cannot be 1, a prime number or 4.*

*Proof.*

Suppose the contrary. Using (3.2) part (4) for the ordinary degree  $(1, 1)$  and the lemmas (3.4) and (3.5) one sees that after a linear change of variables,  $f_{1,1}^+ = \ominus x^s y^t$  with  $s < t$ .

We will deduce that either our Theorem holds or by suitable automorphisms, if necessary, we can arrange that for some positive integers  $u, v$

$$f_{1,1}^+ = \ominus (x^u y^v)^R \text{ and } g_{1,1}^+ = \ominus (x^u y^v)^S,$$

where  $\text{GCD}(R, S) = \text{GCD}(u, v) = 1$ . We will then deduce a final contradiction from this situation to finish the proof.

First we show that the case  $s = 0$  can be reduced to the above situation.

If  $s = 0$ , then, the Newton polygon  $N(f)$  has a side starting with  $(0, t)$  corresponding to a degree  $(a, b)$  with  $a > b > 0$ . Reasoning as before, we find  $b = 1$  and  $f_{a,b}^+ = \ominus ((x + \ominus y^a)^u y^v)^R$  for some positive integer  $R$ . Similarly,  $g_{a,b}^+ = \ominus ((x + \ominus y^a)^u y^v)^S$  for some positive integer  $S$ . Hence  $(v + au)R = n$  and  $(v + au)S = m$ . So  $(v + au)$  divides  $\text{GCD}(n, m)$ , in other words,  $(v + au)$  divides a prime number or 4.

There are only two possibilities:

- (i)  $u, v$  are nonzero,  $\text{GCD}(u, v) = 1$  and  $\text{GCD}(R, S)$  equals 1 or 2
- (ii)  $a = u = 2, v = 0$  and  $\text{GCD}(R, S) = 1$  (this occurs only when  $\text{GCD}(n, m) = 4$ ).

We can clearly apply a suitable automorphism of the form:  $x \rightarrow x + \ominus y^a, y \rightarrow y$  so that the  $y$ -degree reduces, but the hypothesis of the Theorem continues to hold. We can also check that after the automorphism, the new  $(1, 1)$  degree forms for  $f, g$  become  $\ominus (x^u y^v)^R, \ominus (x^u y^v)^S$  respectively. In the case (ii), the new  $\text{GCD}(n, m)$  becomes 2 and we can start the proof again with the assurance that we will not run into case (ii) again. Thus we have achieved the promised reduction.

Now we assume  $s > 0$  and again reduce to the situation mentioned at the beginning of the proof.

Consider the side of  $N(f)$  which starts with  $(s, t)$  and is directed towards the  $x$ -axis. For weights corresponding to this line, we must have  $a > 0, a + b > 0$  and  $b \leq 0$ . As before we write  $f_{a,b}^+ = \ominus C^R$  and  $g_{a,b}^+ = \ominus C^S$  where  $C$  is  $(a, b)$ -homogeneous and *not* a monomial. Further  $\ominus (C_{1,1}^+)^R = f_{1,1}^+$  and  $\ominus (C_{1,1}^+)^S = g_{1,1}^+$ . Write  $C_{1,1}^+ = x^u y^v$ ,  $0 < u < v$ .

Since  $(u + v)R = n$  and  $(u + v)S = m$  and  $\text{GCD}(n, m) = 1$ , prime or 4, we find again that  $\text{GCD}(u, v) = 1$ .



Thus, again, we have achieved the promised reduction.

Now we deduce the final contradiction.

Consider the equation  $dC \wedge dD = Cdx \wedge dy$  with respect to the  $(1,1)$ -degree. If  $\Delta_{1,1}(D) \leq 2$ , then according to (3.5) we find that  $C$  is a monomial in  $x$  and  $y$ .

If  $\Delta_{1,1}(D) > 2$ , then  $dC_{1,1}^+ \wedge dD_{1,1}^+ = 0$ . Hence a power of  $D_{1,1}^+$  is equal to a power of  $C_{1,1}^+$ . But since  $\text{GCD}(u, v) = 1$ , we have in fact,  $D_{1,1}^+ = \vartheta(C_{1,1}^+)^p$  for some  $p$ . Replace now  $D$  by  $D^* = D - \vartheta C^p$ . Then  $dC \wedge dD^* = Cdx \wedge dy$ ,  $D^*$  is  $(a, b)$ -homogeneous and  $\Delta_{1,1}(D^*) < \Delta_{1,1}(D)$ . So finally one finds a  $\hat{D}$  with  $\Delta_{1,1}(\hat{D}) \leq 2$  and  $dC \wedge d\hat{D} = Cdx \wedge dy$ . This implies again the contradiction " $C$  is a monomial".

#### 4. Some interesting calculations for plane curves

*Preamble.* We continue to use the previous notation. Explicitly, we fix polynomials  $f, g$  in  $k[x, y]$ , satisfying the Jacobi-condition. Without loss of generality, we may assume that  $f, g$  are monic in  $y$ . In fact, the best way of describing our setup is to start with the situation as deduced in the beginning of the proof of (3.6), where  $f_{1,1}^+ = (x^s y^t)$  with  $s < t$  and make a linear change  $x \rightarrow y + x$ . Even this change is not quite necessary but avoids technical complications.

Set  $n$  to be the  $y$ -degree of  $f$  and  $m$  to be the  $y$ -degree of  $g$ .

Now the first corner of the Newton diagram (figure 3) will be along the line joining  $(s, t)$  to the origin.

We may, in this case, view the pair  $f, g$  as describing a parametric polynomial plane curve with parameter  $y$  over the field  $k(x)$ . We, therefore get a standard meromorphic Newton-Puiseux expansion of  $g$  in  $\overline{k(x)}((\eta))$  where  $\eta$  is defined by  $f = \eta^{-n}$  and  $\overline{k(x)}$  denotes the algebraic closure of  $k(x)$ . Using the change of variables from  $x, y$  to  $x, \eta$  we compute  $J_{x,\eta}(f, g)$  or the  $x$ -derivative of the  $\eta$ -expansion. Then it is easy to see that we have:

$$g = \eta^{-m} + \dots + (\vartheta x + c)\eta^{n-1} + \dots \in k[x]((\eta)).$$

In other words, the Newton-Puiseux expansion of  $g$ , indeed lives over the field  $k(x)$ . Moreover, all the terms of the expansion up to the displayed term of order  $n-1$  are free of  $x$ .

Taylor Resultant Theorem 4.1. [A2, P. 153]. *Given any two nonconstant polynomials*

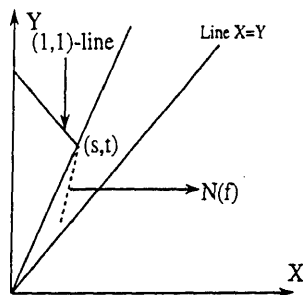


Figure 3.

$p(t), q(t)$  over a field  $K$  of characteristic 0, set

$$\Theta(t) = \text{Res}_\tau \left( \frac{p(t) - p(\tau)}{t - \tau}, \frac{q(t) - q(\tau)}{t - \tau} \right).$$

Set  $A = K[p(t), q(t)]$  and let  $\mathfrak{C}$  be the conductor of the ring  $A$  in the ring  $A' = K[t]$ . Then  $\mathfrak{C} = \Theta(t)A'$ . In particular, thinking of  $p(t), q(t)$  as parametrizing the plane curve

$$H(X, Y) = \text{Res}_t(p(t) - X, q(t) - Y)$$

we get:

- (1) The parametrization is faithful if and only if  $K(t) = \text{Qt}(A)$  or, equivalently,  $\Theta(t) \neq 0$ .
- (2) If  $\Theta \neq 0$ , then the curve is nonsingular (at finite distance) if and only if  $A = A'$  or, in other words,  $\Theta = \vartheta \in K$ .

*Proof.* Let  $p(t)$  have degree  $n$  and set  $E = \text{Qt}(K[p(t)])$ . Then  $K(t)$  is a separable algebraic extension of degree  $n$  over  $E$ . Let  $t_1, \dots, t_n = t$  be a full set of conjugates of  $t$  in some fixed algebraic closure of  $E$ . Note that we have the factorizations:

$$\frac{p(t) - p(\tau)}{t - \tau} = \prod_{1 \dots (n-1)} (\tau - t_i).$$

By the defining property of resultants, we get that

$$\Theta(t) = \vartheta \prod_{1 \dots (n-1)} \frac{q(t) - q(t_i)}{t - t_i}$$

where the nonzero constant is needed to handle signs and powers of the leading coefficient if any.

We also have from the definition of  $H(X, Y)$  that

$$H(p(t), Y) = (-1)^n \prod_{1 \dots n} (Y - q(t_i))$$

and so

$$H_Y(p(t), q(t)) = (-1)^n \prod_{1 \dots (n-1)} (q(t) - q(t_i)).$$

Also, if we pick a new indeterminate  $W$ , then

$$p(W) - p(t) = \prod_{1 \dots n} (W - t_i), \quad \text{so } p'(t) = \prod_{1 \dots (n-1)} (t - t_i).$$

From the above calculations, it follows that

$$\Theta(t) = \vartheta \frac{H_Y(p(t), q(t))}{p'(t)}.$$

The Dedekind Formula for the conductor and the different exactly says that the

fractionary ideal generated by the right hand side of the above equation is the conductor ideal  $\mathfrak{C}$ . This establishes the main formula.

The parametrization is faithful if and only if the field  $K(t)$  coincides with  $Q_t(A)$ . Now the fields  $K(t)$  and  $Q_t(A)$  are distinct if and only if

$$[Q_t(A):K(p(t))] < n = [K(t):K(p(t))]$$

or in other words  $q(t)$  has less than  $n$  conjugates over  $K(p(t))$ . But this last condition is equivalent to  $q(t) = q(t_i)$  for some  $i < n$ , or, in other words, to  $\Theta(t) = 0$ .

The rest of the proof follows from the properties of the conductor.

*Semigroups.* We now present some connections between Newton-Puiseux expansions and certain value-semigroups associated with plane curves.

We begin with a field  $K$  whose characteristic does not divide a given integer  $M_0$ . We consider a (Laurent) power series

$$\phi(\eta) = \sum \alpha_i \eta^i \in K((\eta))$$

Without considering the genesis of this series, we build a set of associated sequences as follows.

First set  $S = \text{supp}(\phi) = \{i | \alpha_i \neq 0\}$ . Set  $M_1$  to be the order of  $\phi$ , i.e.  $\text{ord}_\eta \phi(\eta) = \inf\{i | i \in S\}$ . Also set  $d_1 = |M_0|$ ,  $d_2 = \text{GCD}(M_1, d_1)$ . Further, set  $n_1 = d_1/d_2$ .

Assume that we have already inductively constructed  $M_1, \dots, M_i$  as well as associated sequences  $q_2, \dots, q_i, d_1, \dots, d_{i+1}$  and  $n_1, \dots, n_i$ . We extend this construction as follows.

Set  $M_{i+1} = \inf\{j > M_i | j \in S \text{ and } j \text{ is not divisible by } d_{i+1}\}$ . In case  $M_{i+1} = \infty$  we set  $h = i$  and declare the process finished. Otherwise, set  $q_{i+1} = M_{i+1} - M_i$ ,  $d_{i+2} = \text{GCD}(M_{i+1}, d_{i+1})$  and  $n_{i+1} = d_{i+1}/d_{i+2}$ .

Of course, the inductive definition is then continued.

There are two more associated sequences which can be now defined in terms of the above.

$$s_0 = M_0 \text{ and } s_i = \sum_1^i q_j d_j \text{ for } 1 \leq i \leq h.$$

Also, we set

$$r_0 = s_0 \text{ and } r_i = s_i/d_i \text{ for } 1 \leq i \leq h.$$

It is easy to see that  $\text{GCD}(r_0, \dots, r_i) = d_{i+1}$  and hence any integer  $b$  is an integral combination of  $r_0, \dots, r_i$  iff  $d_{i+1}$  divides  $b$ .

*Strict generation.* A combination  $\sum a_i r_i$  is said to be a **strict combination** if

- (1)  $a_0 \geq 0$  and
- (2)  $0 \leq a_i < n_i$  for  $1 \leq i \leq h$ .

We say that  $\{r_i\}$  form a **strict set of generators** if the semigroup generated by  $r_0, \dots, r_i$  consists entirely of their strict combinations.

There are two special cases of interest when we get a strict set of generators  $\{r_i\}$  from the given  $\phi$  and  $M_0$ . See [A1, Chapter 8].

*Algebroid Case:* Here  $M_0 > 0$ ,  $x = \eta^{M_0}$  and  $y = \phi(\eta)$  describe the local Newton-Puiseux expansion for a plane curve at a point in the plane, provided the curve has only one branch at the point. In this case,  $M_1 > 0$  and  $\phi(\eta) \in K[[\eta]]$ . If  $A$  denotes the local ring of the point on the curve, then the unique valuation of the curve centered at  $A$  coincides with the  $\eta$ -adic order and the semigroup of values of all the (nonzero) elements of  $A$  is the semigroup (strictly) generated by  $\{r_i\}$ .

*Meromorphic Case:* This is a special case of a meromorphic curve. Here,  $M_0 < 0$ ,  $x = \eta^{M_0}$  and  $y = \phi(\eta)$  describe the expansion at infinity for a plane curve, provided the curve has only one branch (or place) at infinity. In this case, all  $r_i$  are negative. If  $A$  denotes the coordinate ring of the curve, then the  $\eta$ -adic order denotes the unique valuation at infinity (nonpositive on  $A$ ) and the semigroup of orders of (nonzero) elements of  $A$  coincides with the semigroup (strictly) generated by  $\{r_i\}$ .

Thus, any properties deducible from strict generation apply to each of these situations.

*Uniqueness of expression.* Assume that  $\{r_i\}$  is any sequence of nonzero integers with  $0 \leq i \leq h$ . Assume that  $d_i = \text{GCD}(r_0, \dots, r_{i-1})$  for  $1 \leq i \leq h+1$  and that  $n_i = d_i/d_{i+1}$  for  $1 \leq i \leq h$ . Let  $\Gamma$  denote the semigroup generated by  $\{r_i\}$ .

If  $b$  is any integer in the group generated by  $\{r_i\}$  or equivalently, if  $d_{h+1}$  divides  $b$ , then we have a unique (partially strict) expression

$$b = \sum_0^h a_i r_i \text{ where } 0 \leq a_i < n_i \text{ for } 1 \leq i \leq h.$$

Moreover, if  $\{r_i\}$  form a strict set of generators for  $\Gamma$ , then  $b \in \Gamma$  iff  $a_0 \geq 0$ .

*Proof.* This is standard stuff as in [A1, Chapter 1]. The main idea is as follows. The remark about the condition for  $b$  to be in the group is obvious from the fact that the GCD of  $r_0, \dots, r_h$  is  $d_{h+1}$ . It is also obvious that  $n_i r_i$  is divisible by  $d_i$  and hence is an integral combination of  $r_0, \dots, r_{i-1}$ . Thus any given integral combination of  $b$  in terms of  $r_0, \dots, r_h$  can easily be transformed to the desired (partially strict) form. Uniqueness is deduced from the GCD properties by induction on the last  $r_i$  present in the expression.

The last assertion follows easily from the definition of strict generation.

*A Symmetry Property.* Assume that  $\{r_i\}$  form a strict set of generators for a semigroup  $\Gamma$ . Set

$$\sigma = -r_0 + \sum_1^h (n_i - 1)r_i$$

Given any two integers  $u, v$  divisible by  $d_{h+1}$ , such that  $u + v = \sigma$ , one and only one of  $u, v$  belongs to  $\Gamma$ .

*Proof.* Clearly  $u, v$  belong to the group generated by  $\{r_i\}$ . Write the unique (partially strict) expression for  $u$  as

$$u = \sum_0^h a_i r_i \text{ where } 0 \leq a_i < n_i \text{ for } 1 \leq i \leq h$$

Then obviously the unique (partially strict) expression for  $v$  is given by

$$v = (-1 - u_0)r_0 + \sum_1^h (n_i - u_i - 1)r_i.$$

Thus, from the above criterion we have:

$$u \in \Gamma \text{ iff } u_0 \geq 0 \text{ and } v \in \Gamma \text{ iff } (-1 - u_0) \geq 0.$$

Clearly among the two integers  $u_0$  and  $-1 - u_0$  exactly one can be nonnegative, hence the result.

*Conductor of a semigroup.* If a semigroup consists of nonnegative integers only. The **conductor of the semigroup is defined to be** the least nonnegative integer such that it and all bigger integers are in the semigroup.

If a semigroup consists of nonpositive integers only, the **conductor of the semigroup is defined to be** the largest nonpositive integer such that it and all smaller integers are in the semigroup.

*A Formula For The Conductor.* Assume that  $\{r_i\}$  form a strict set of generators for a semigroup  $\Gamma$ . To simplify notation, further assume that  $d_{h+1} = 1$ , i.e. that the group generated by  $\{r_i\}$  consists of all integers.

Set as before:

$$\sigma = -r_0 + \sum_1^h (n_i - 1)r_i$$

If  $\Gamma$  consists entirely of nonnegative integers, then the number  $c = \sigma + 1$  has the properties:

- (1)  $c$  is an even integer.
- (2) Every integer bigger than or equal to  $c$  is in  $\Gamma$  and  $c$  is the smallest integer with this property. In other words,  $c$  is the conductor of  $\Gamma$ .
- (3) There are exactly  $c/2$  positive integers not in  $\Gamma$ .

If  $\Gamma$  consists entirely of nonpositive integers, then the number  $c = \sigma - 1$  has the properties:

- (1)  $c$  is an even integer.
- (2) Every integer smaller than or equal to  $c$  is in  $\Gamma$  and  $c$  is the largest integer with this property. In other words  $c$  is the conductor of  $\Gamma$ .
- (3) There are exactly  $c/2$  negative integers not in  $\Gamma$ .

*Proof.* Suppose first that  $\Gamma$  consists of nonnegative integers only. By assumption  $-b \notin \Gamma$  if  $b > 0$  and hence by the symmetry property,  $\sigma - (-b) = \sigma + b \in \Gamma$ .

Again by the symmetry property,  $0 \in \Gamma$  and hence  $\sigma - 0 = \sigma \notin \Gamma$ . Thus, no number smaller than  $c$  has the desired property.

Also,  $\sigma$  must be odd, since if  $\sigma = 2b$ , then  $b = \sigma - b$  will be both in  $\Gamma$  as well as

*Length of the integral closure: Algebroid case.* Assume that we have the Algebroid Case described above and  $A$  is the local ring of a point on the curve. Let  $A'$  be the integral closure of  $A$  in its quotient field (the function field of the curve) and let  $\mathfrak{C}$  be the conductor of  $A'$  over  $A$ . Let  $\Gamma$  be the semigroup of values of (nonzero) elements of  $A$  in the unique valuation centered at the point and let  $c$  be the conductor of  $\Gamma$  as described above. Then the length of  $A'/A$  as an  $A$ -module is exactly  $c/2$ . By the well known Gorenstein property of such local rings, the length also coincides with the length of  $A/\mathfrak{C}$  as an  $A$ -module. In particular, the length is described by the formula

$$\frac{1 - r_0 + \sum_1^h (n_i - 1)r_i}{2}.$$

This verifies the formula on page 169 of [A2]

*Proof.* To simplify matters, we assume that the field  $K$  is algebraically closed. It is clear that if we take the values of various nonzero elements of  $A'$  we get the semigroup of all nonnegative integers. List the  $c/2$  nonnegative integers which do not belong to  $\Gamma$  as  $u_1, \dots, u_{c/2}$  and pick a sequence of  $c/2$  elements of  $A'$  with values  $u_1, \dots, u_{c/2}$  respectively. It is easy to see that they form a basis of  $A'/A$  over  $K$  and in fact determine the length of  $A'/A$ .

The proof for the general  $K$  should be carried out by the already mentioned Dedekind formula or by the technique of extending the ground field. We omit these technical details.

*Length of the integral closure: Meromorphic case.* Assume that we have the Meromorphic case described above and  $A = K[x, y]$  is the affine coordinate ring of a curve having one place at infinity. Let  $A'$  be the integral closure of  $A$  in its quotient field (the function field of the curve) and let  $\mathfrak{C}$  be the conductor of  $A'$  over  $A$ . Let  $\Gamma$  be the semigroup of values of (nonzero) elements of  $A$  in the unique valuation at infinity, i.e., the unique valuation of the function field not containing  $A$ . Let  $c$  be the conductor of  $\Gamma$  as described above. Then the length of  $A'/A$  as an  $A$ -module is exactly  $c/2$ . By the well-known Gorenstein property of such rings, the length also coincides with the length of  $A/\mathfrak{C}$  as an  $A$ -module. In particular, the length is described by the formula

$$\frac{1 - r_0 + \sum_1^h (n_i - 1)r_i}{2}.$$

*Proof.* The proof is formally the same as in the Algebroid case.

## 5. The Newton Puiseux expansions for different weights

*Preamble.* We wish to generalize the Newton-Puiseux expansions discussed in the previous section by making the expansion which will respect a certain weight.

We assume that  $f, g$  are polynomials in  $x, y$  of degree at least 2 satisfying the Jacobi-condition.

Fix rational weights  $a, b$  and set  $w = (a, b)$ . Assume that  $\Delta_w(f) > 0$  and  $\Delta_w(g) > 0$ . Moreover, we are generally interested in weights corresponding to Newton lines only,

least one of  $a, b$  is positive and since weights proportional by positive numbers yield the same degree forms, we may assume  $a, b$  to be coprime integers. By interchanging  $x, y$  if necessary, we are reduced to considering the cases  $a = 0, b = 1$  or  $a > 0$  with  $\text{GCD}(a, b) = 1$ . The case when  $a = 0$  is the case of considering the  $y$ -degree as the weight and the corresponding expansion is as discussed in the last section. We therefore assume that  $a > 0$ . We also make the following assumption, which is necessary for the validity of some of the technical results.

**Special assumption:** Assume that the weights  $a, b$  are such that  $a + b > 0$  or that the weight of the monomial  $xy$  is positive.

Generally, this assumption is valid for weights along Newton lines starting above the line  $X = Y$ , since in the contrary case, the resulting Newton diagram will not cross the  $X = Y$  line and the resulting polynomials  $f, g$  will be divisible by  $y$  and the Jacobi-condition fails. In our current set up, this will be true for the sequence of Newton lines starting from the end of the  $(1, 1)$ -line until the first line which crosses the  $X = Y$  line.

We now set  $y = zt^b, x = t^a$ . Note that the change of variables from  $(x, y)$  to  $(z, t)$  causes the Jacobian to be multiplied by  $\ominus t^{(a+b-1)}$ .

We now think of  $f, g$  as elements in the field  $k(z)((\tau))$ , the field of meromorphic power series in  $\tau = t^{-1}$  over the field of rational functions in  $z$ . Note that the weight of any of the original polynomials can simply be read off as the highest power of  $t$  occurring, or, equivalently, the negative of the  $\tau$ -order.

The change of variables to  $(z, \tau)$  causes the original Jacobian to be multiplied by  $\ominus t^{(a+b-1)} \tau^{-2} = \ominus \tau^{-(a+b+1)}$ .

In particular, we can write  $f_w^+ = \ominus t^{\Delta_w(f)} P(z)$  where  $P(z)$  is some polynomial. Consequently,

$$f = \ominus P(z) t^{\Delta_w(f)} P^*(z, \tau) = \ominus P(z) \tau^{-\Delta_w(f)} P^*(z, \tau)$$

where  $P^*(z, \tau)$  is a polynomial in  $\tau$  with coefficients in  $k(z)$ , thought of as an element of  $k(z)[[\tau]]$  and in fact, it is a unit in the power series ring  $k(z)[[\tau]]$ . By taking its  $-\Delta_w(f)$ -th root, we can write

$$f = P(z) \eta^{-\Delta_w(f)}$$

where  $\eta = \tau(P^*(z, \tau))^{-1/(\Delta_w(f))}$  is a new generator for the power series ring  $k(z)((\tau))$  over  $k(z)$ .

Thus, the transformation from  $(x, y)$  to the new variables  $(z, \eta)$  multiplies the Jacobian by a unit times  $\eta^{-(a+b+1)}$ . Moreover, the  $w$ -degree can be computed as the negative of the  $\eta$ -order.

Given any power series

$$G = \sum a_i(z) \eta^i$$

we can compute its Jacobian with  $f$  to be

$$J_{z, \eta}(f, G) = \sum_i (iP'(z)a_i + \Delta_w(f)P(z)a'_i) \eta^{i-1-\Delta_w(f)}.$$

In particular, if  $G$  is obtained from a polynomial in  $(x, y)$  by the change of variables

explained above, we can check that  $f, G$  are  $w$ -related if and only if the term

$$-P'(z)Q(z)\Delta_w(G) + \Delta_w(f)P(z)Q'(z) \quad (*)$$

equals 0, where the leading term of  $G$  is written as  $Q(z)\eta^{-\Delta_w(G)}$ .

Moreover, in the unrelated case, the expression  $(*)$  coincides with a nonzero constant times the leading coefficient of the transformation of the usual jacobian of  $f, G$ .

If  $f, G$  are  $w$ -related, then solving the differential equation obtained by equating  $(*)$  to 0, we deduce that

$$P^{\Delta_w(G)} = -\Theta Q^{\Delta_w(f)}.$$

Let us fix a polynomial  $H = H(z)$  such that we can write for some positive integers  $\nu, \delta'_0$ :

$$f = (H\eta^{-\nu})^{\delta'_0}$$

and such that this kind of expression does not hold for any polynomial of degree smaller than that of  $H(z)$ . Let us denote the expression  $H\eta^{-\nu}$  by  $\zeta$ .

Then the above relatedness condition gets replaced by " $Q(z)\eta^{\Delta_w(G)}$  is an integral power of  $\zeta$ ".

Thus, any power series  $G$  as above can be split in three parts:

$$G = \text{terms involving powers of } \zeta + a\eta^s + \text{higher terms}$$

where  $a\eta^s$  is the first term unrelated with  $f$ . In particular, we have

$$-\Delta_w(f) + s = 1 + \text{ord}_\eta(J_{z,\eta}(f, G))$$

*Newton-Puiseux expansions for a given weight.* 5.1 Applying the above substitutions to  $g$ , we see that it develops into a power series:

$$g = \sum a_i(z)\eta^i$$

We wish to set up the usual characteristic sequence associated with it, as commonly done in the expansion techniques.

Begin by setting  $M_1$  to be the  $\eta$ -order of  $g$  and set  $d_1 = \Delta_w(f)$  or the negative of the  $\eta$ -order of  $f$ . Set  $d_2 = \text{GCD}(M_1, d_1)$  and  $n_1 = d_1/d_2$ .

Assume that we have inductively defined  $M_1, \dots, M_i$  along with  $d_1, \dots, d_{i+1}$ ,  $q_2, \dots, q_i$  and  $n_1, \dots, n_i$ . Then we set  $M_{i+1}$  to be the first exponent of  $\eta$  in the support of the expansion of  $g$ , which is not divisible by  $d_{i+1}$ . Set  $d_{i+2} = \text{GCD}(M_{i+1}, d_{i+1})$ ,  $n_{i+1} = d_{i+1}/d_{i+2}$  and  $q_{i+1} = M_{i+1} - M_i$ . In case there is no such exponent, we declare the process finished and set  $h = i$ . We say that we have  $h$  characteristic pairs  $(M_i, d_i)$ .

Strictly speaking, this whole construction depends on the choice of  $w$ , but we have chosen not to clutter up the notation by tacking on an extra subscript.

We can visually display the characteristic sequence by writing:

$$g = c_1\eta^{M_1} + \dots + c_2\eta^{M_2} + \dots + c_h\eta^{M_h} + \dots$$

Associated with the above sequence is the sequence of "pseudoapproximate" roots, which are certain polynomials in  $f, g$ , which we now introduce.



Set  $g_0 = f$  and  $g_1 = g$ . Define  $\delta_0 = \Delta_w(f)$ ,  $\delta_1 = -M_1 = \Delta_w(g)$ . Also, set  $\mu_i = \delta_0 + \delta_1 - \sum_{j=2}^i q_j$  for  $i = 1, \dots, h$ . For technical reasons, set  $\mu_0 = \infty$ . Note that  $\mu_i$  is also equal to  $\delta_0 - M_i$ . Also set  $\delta_i = n_{i-1} \delta_{i-1} - q_i$ . It is a standard calculation to check the identity (for  $1 \leq i \leq h$ ):

$$\mu_i = \delta_0 + \delta_i - \sum_{j=1}^{i-1} (n_j - 1) \delta_j.$$

Note that by the known transformations, the Jacobian of  $f, g$  with respect to  $z, \eta$  has  $\eta$ -order  $-a - b - 1$  and so the first unrelated term in the expansion of  $g = g_1$  has  $\eta$  order  $-a - b + \delta_0$ . Thus, we can write

$$g = g_1 = \eta^{-\delta_1} (c_1 + \dots + c \eta^{\delta_0 + \delta_1 - a - b} + \dots)$$

where  $c$  is a nonzero polynomial in  $z$  and all earlier terms are related to  $f$ . In particular,

$$g_1 = \eta^{-\delta_1} (c_1 + \dots + c \eta^{\mu_1 - a - b} + \dots).$$

Now suppose that we have inductively built  $g_0, \dots, g_i$  such that for  $1 \leq j \leq i$ :

- (1) The  $\eta$ -order of  $g_j$  is  $-\delta_j$ , so that its  $w$ -weight is precisely  $\delta_j$ .
- (2) The first term in the expansion of  $g_j$  which is unrelated to  $f$  has  $\eta$ -order  $\mu_j - a - b - \delta_j$ .
- (3)  $g_j$  is related to  $f$ .

Then we try to build  $g_{i+1}$  as follows.

By a **standard monomial** in  $g_0, \dots, g_i$  we mean a monomial of the form  $\prod_j g_j^{\alpha_j}$  where  $0 \leq \alpha_j < n_j$  for  $1 \leq j \leq i$ , while  $0 \leq \alpha_0$ . We will conveniently shorten the notation to write  $g^\alpha$  for  $\prod_j g_j^{\alpha_j}$ .

For any  $1 \leq j \leq i$  the  $\eta$ -order of  $g_j$  is  $-\delta_j$ , while the  $\eta$ -order of the first term in  $g_j$  unrelated to  $f$  is  $\mu_j - \delta_j - a - b$ . Thus,  $g_j$  is related to  $f$  if and only if this term is not the leading term of  $g_j$ , i.e.  $-\delta_j < \mu_j - \delta_j - a - b$ , or equivalently,  $a + b < \mu_j$ . We know this for  $1 \leq j \leq i$  already.

We begin by a trial value  $v = g_i^{\alpha_i}$ .

We keep on modifying  $v$  until it becomes  $g_{i+1}$ .

- (1) If the  $w$ -leading form of  $v$  cannot be expressed as the  $w$ -leading form of  $ag^\alpha$  for some  $a \in k$  and some standard monomial  $g^\alpha$  in  $g_0, \dots, g_i$ , then we declare  $g_{i+1} = v$  and stop this modification.
- (2) If there is a standard monomial  $g^\alpha$  in  $g_0, \dots, g_i$  such that  $v$  has the same  $w$ -leading form as  $cg^\alpha$  for some  $c \in k$ , then we modify  $v$  to  $v - cg^\alpha$ . Note that the  $w$ -degree of  $v$  decreases in this process and so the modification has to eventually stop.

We need to verify that  $g_{i+1}$  has the correct order and indeed that it is the next "pseudoapproximate root" as in the expansion techniques.

For  $f = g_0$ , we might consider the difference between the leading term and the first term unrelated to  $f$  to be  $\infty$ , and hence equals  $\mu_0 - a - b = \infty$ .

Now, in the buildup of  $g_{i+1}$  described above, the first step of cancelling a monomial causes this "gap of the first unrelated term" to decrease from  $\mu_i - a - b$ . Any subsequent modifications only increase the  $\eta$ -order without affecting the first unrelated term, since the modifying terms now all have bigger gaps.

Thus the final gap for  $g_{i+1}$  is  $\mu_i - a - b - q$  for some  $q$ . We wish to show that  $q = q_{i+1}$ .

By the standard theory,  $q \leq q_{i+1}$  since the modification cannot be pushed beyond that even with coefficients in  $k(z)$ . If  $q < q_{i+1}$ , then the  $\eta$  order of our  $g_{i+1}$  is in the semigroup generated by the  $\eta$ -orders of  $g_0, \dots, g_i$ . The fact that our modification process stopped means that the multiplier coefficient needed is not in  $k$ . It is then evident that the highest  $z$ -degree term in  $g_{i+1}$  cannot be related to that of  $f$ . From the remarks in (3.5) and our special assumptions, it follows that the leading form of  $f$  must be a monomial. Since we have also assumed that we have weights whose degree form is a line, we are done.

Clearly, this process then continues, until we reach  $h$  or we reach an unrelated  $g_i$ . If  $i = h$  and  $g_h$  is still related to  $f$ , then we can continue the modification until we reach an unrelated  $g_{h+1}$ . However, it cannot correspond to any characteristic term (since we have gone past all such terms) and we can deduce that the leading form of  $f$  must have been a monomial. We summarize this in:

*Pseudoapproximate roots along a Newton Line 5.2: Assume that we have weight  $w = (a, b)$  with  $a > 0$  and  $a + b > 0$  such that the degree form of  $f$  is not a monomial. Then, in the above notation, there is a sequence of pseudoapproximate roots  $g_0, \dots, g_i$  for some  $i \leq h$  such that each  $g_j$  is a polynomial in  $f, g$  over  $k$ . Moreover,  $g_i$  is  $w$ -unrelated to  $f$  and we have  $\mu_i = a + b$ .*

*Definition.* Let  $w$  be any rational weight such that either  $w = (0, 1)$  or  $w = (a, b)$  with  $a > 0$ . We will say that  $f$  has  $i$  pseudoapproximate roots along  $w$ , if we can construct the sequence  $g_0, \dots, g_i$  as described above. Note that the number  $i$  can be smaller than the usual number  $h$  given by the expansion techniques and can even be  $h + 1$  when the degree form relative to  $w$  is a monomial.

Now we consider the variation of the number of pseudoapproximate roots corresponding to two consecutive Newton Lines. Let  $w_1 = (a_1, b_1)$  and  $w_2 = (a_2, b_2)$  be the consecutive weights, so that  $b_2/a_2 < b_1/a_1$ . Also assume that the common corner for the Newton diagram of  $f$  is a point  $(s_1, s_2)$  above the line  $X = Y$  (i.e.  $s_1 < s_2$ ). The point corresponds to the lowest  $y$ -degree term for  $f_{w_1}^+$  and the highest  $y$ -degree term for  $f_{w_2}^+$ . Let the sequence  $g_1, \dots, g_i$  be constructed for the weight  $w_1$ , as shown above.

Then,  $\Delta_{w_1}(f) = a_1 s_1 + b_1 s_2$  and  $\Delta_{w_2}(f) = a_2 s_1 + b_2 s_2$ . Set

$$\lambda = \frac{a_2 s_1 + b_2 s_2}{a_1 s_1 + b_1 s_2}$$

and

$$\lambda^* = \frac{a_2 + b_2}{a_1 + b_1}.$$

Recall that  $g_j$  is  $w_1$ -related to  $f$  if and only if  $\mu_j/(a_1 + b_1) > 1$  and, in fact, we have:

$$\frac{\mu_1}{a_1 + b_1} > \frac{\mu_2}{a_1 + b_1} > \dots > \frac{\mu_i}{a_1 + b_1} = 1.$$

From the alternate formula for  $\mu_j$ , it is clear that  $\mu_j$  corresponding to  $w_2$  is  $\lambda\mu_j$ . To see if  $g_j$  is also  $w_2$ -related to  $f$ , we need to check if  $\lambda\mu_j/(\dot{a}_2 + b_2) > 1$  or

$$\frac{\mu_j}{a_1 + b_1} > \lambda^*/\lambda.$$

With our hypothesis, it is easy to check that  $\lambda^*/\lambda > 1$  and so the condition for relatedness gets tighter as we move from  $w_1$  to  $w_2$ . Indeed, if  $i^*$  is the last pseudoapproximate root for the weight  $w_2$ , then we must have,

$$\frac{\lambda\mu_1}{a_2 + b_2} > \frac{\lambda\mu_2}{a_2 + b_2} > \dots > \frac{\lambda\mu_{i^*}}{a_2 + b_2} = 1$$

or, in other words:

$$\frac{\lambda}{\lambda^*} \frac{\mu_1}{a_1 + b_1} > \frac{\lambda}{\lambda^*} \frac{\mu_2}{a_1 + b_1} > \dots > \frac{\lambda}{\lambda^*} \frac{\mu_{i^*}}{a_1 + b_1} = 1.$$

Thus, we have  $\mu_{i^*} = (a_1 + b_1)\lambda^*/\lambda$ . Naturally,  $i^* < i$ . Consider the possibility that  $i^* < i - 1$ . We choose a weight  $w_3 = (a_3, b_3)$  such that

$$\frac{a_3 s_1 + b_3 s_2}{a_1 s_1 + b_1 s_2} \frac{\mu_{i-1}}{a_3 + b_3} = 1$$

or, in other words,

$$\frac{\mu_{i-1}}{a_1 + b_1} = \frac{a_1 s_1 + b_1 s_2}{a_3 s_1 + b_3 s_2} \frac{a_3 + b_3}{a_1 + b_1}.$$

Using (the consequence of the special assumption)  $a_1 + b_1 > 0$ , it is easy to verify that the expression  $\psi(a, b) = \frac{a_1 s_1 + b_1 s_2}{a s_1 + b s_2} \frac{a + b}{a_1 + b_1}$  is a monotonic decreasing function of the ratio  $b/a$  and consequently, the weight line corresponding to  $w_3$  lies between the lines for  $w_1, w_2$ . Since, the  $w_1, w_2$  lines were assumed consecutive, this implies that the degree form of  $f$  for the weight  $w_3$  must be the same monomial  $\ominus x^{s_1} y^{s_2}$ . Also, clearly, for  $w_3$  we have exactly one less pseudoapproximate root than for  $w_1$ .

What we have obtained is the result:

*Variation along Newton lines 5.3. Assume that two consecutive Newton Lines of  $f$  share a common vertex and let  $x^{s_1} y^{s_2}$  be the monomial pointing towards the vertex, chosen as described above. Further assume that:*

- (1) The two weights  $w_1 = (a_1, b_1)$  and  $w_2 = (a_2, b_2)$  are such that  $a_1, a_2$  are positive.
- (2)  $b_2/\dot{a}_2 < b_1/a_1$ .
- (3)  $s_1 < s_2$ .

Then  $f$  has at least one less pseudoapproximate root along  $w_2$  than along  $w_1$ . Moreover, there exists an intermediate weight  $w_3 = (a_3, b_3)$  with

*Note.* Note that we are discussing the concept of pseudoapproximate roots corresponding to a given weight and we quit developing the pseudoapproximate roots as soon as we reach an unrelated root corresponding to the weight. Thus, as we start with the starting weight  $(0, 1)$  and consider various values  $(a, b)$  with the ratio  $b/a$  steadily decreasing, we march along the Newton diagram. What we have shown here is that we keep on getting fewer and fewer pseudoapproximate roots related to  $f$  until we cross the line  $X = Y$ . Afterwards, generally the function  $\psi(a, b)$  turns increasing and the number of roots tends to increase. In fact, the sign of the derivative of the function  $\psi(a, b)$  is determined by the sing of  $(s_1 - s_2)/(a + b)$  and after crossing the line  $X = Y$  the sign turns positive, unless the special assumption also fails and  $a + b$  turns negative.

It is possible to make an independent argument to show that the highest  $y$ -degree term of the unrelated pseudoapproximate root must become unrelated to that of  $f$  below the  $X = Y$  line. Thus, in view of the remarks in (3.5), the special assumption must fail after crossing the  $X = Y$  line. The analysis of Newton Lines in this region after the  $X = Y$  line is not relevant for the remaining part and hence no further discussion is provided here.

**Lemma 5.4. Lower bound on the number of pseudoapproximate roots.** *Let a weight  $w = (a, b)$  corresponding to a Newton Line of  $f$  satisfy  $0 < a$  and  $b < a$ . Then  $f, g$  must be  $w$ -related. In other words, the number of pseudoapproximate roots for the weight  $w$  is at least 2.*

*Proof.* This is only a special case of the proof in the beginning of (3.2) where the result is proved when either  $a$  or  $b$  are positive.

**Case of two characteristic terms 5.5.** *Assume that  $f$  and  $g$  have at most two characteristic terms for the  $(1, 1)$  weight and satisfy the rest of the conditions described in the preamble. Then the Jacobian theorem holds for  $f, g$ .*

*Proof.* In view of the various results from the earlier sections, it is clear that we are reduced to considering the case where the first corners after the  $(1, 1)$  weight line for  $f, g$  are of the form  $(pt_1, pt_2), (qt_1, qt_2)$  respectively, where  $0 < t_1 < t_2$  and  $p, q$  are coprime. Let  $w = (a, b)$  be the weight for the next line. Clearly, we have  $0 < a$  and  $b < a$ .

By Lemma 5.3, we can have at most one pseudoapproximate root along  $w$ . This means  $f, g$  must be  $w$ -unrelated.

On the other hand by Lemma 5.4, we get that  $f, g$  must be  $w$ -related. This is a contradiction.

## References

- [A1] Abhyankar S S, Expansion Techniques in Algebraic Geometry, Tata Institute of Fundamental Research, Bombay, 1977
- [A2] Abhyankar S S, Algebraic geometry for Scientists and Engineers, *Mathematical Surveys* (35) AMS., 1990
- [Ca] Campbell L A, A Condition For a Polynomial To Be Invertible, *Math. Ann.* **205** (1973), 243–248
- [Ch] Chevalley C, Algebraic Functions of One variable, *Mathematical Surveys* (6) AMS, 1951
- [Ke] Keller O, Ganze Cremona Transformationen, *Mh. Math. Phys.* **47** (1993), 299–306
- [Na] Nagata M, *Local Rings*, John Wiley, New York, 1962
- [Ra] Raynaud M, *Anneaux Locaux Henseliens*, Springer Verlag Lecture Notes in Mathematics (169). 1939

# On polynomial isotopy of knot-types

RAMA SHUKLA

Department of Mathematics, Indian Institute of Technology, Bombay 400 076, India  
Present address: Mehta Research Institute of Mathematics and Mathematical Physics,  
Allahabad 211 002, India

MS received 3 February 1993; revised 28 September 1993

**Abstract.** We have proved that every knot-type  $\mathbb{R}^3, \mathbb{R}^3$  can be uniquely represented by polynomials up to polynomial isotopy i.e. if two polynomial embeddings of  $\mathbb{R}$  in  $\mathbb{R}^3$  represent the same knot-type, then we can join them by polynomial embeddings.

**Keywords.** Non-compact knots; polynomial isotopy.

## 1. Introduction

Intuitively by a knot we mean a simple closed curve in  $\mathbb{R}^3$  which is not the boundary of a smoothly embedded disc in  $\mathbb{R}^3$ , otherwise it is a *trivial knot*. Mathematically we define a knot as a smooth embedding of  $S^1$  in  $\mathbb{R}^3$  (or equivalently as a smooth embedding of  $S^1$  in  $S^3$ ). We identify a knot by its image and sometimes say that  $K$  is a knot, which means that  $K$  is the image of a smooth embedding of  $S^1$  in  $\mathbb{R}^3$ . Two knots  $K_1$  and  $K_2$  given by embeddings  $\phi_0$  and  $\phi_1$  of  $S^1$  in  $S^3$  are called ambient isotopic if there exists an orientation preserving diffeomorphism  $h: S^3 \rightarrow S^3$  such that  $h(K_1) = K_2$ , or equivalently there exists an isotopy  $F: S^1 \times I \rightarrow S^3$  between  $\phi_0$  and  $\phi_1$ .

Let  $K$  be a knot given by the embedding  $\phi \equiv (\alpha, \beta, \gamma): S^1 \rightarrow \mathbb{R}^3$ . In order to work with  $K$  we project it into a *suitable plane*. By suitable we mean that there are only finitely many *singular points* in the projected image and all of them are *ordinary double points* (i.e. the embedding followed by the projection is a *generic immersion*). Such a projection is called a *regular projection*. Suppose that  $K$  has a regular projection on the  $xy$  plane. Then for each double point in the projection there exist  $t_1, t_2 \in S^1$  ( $t_1 \neq t_2$ ) such that  $(\alpha(t_1), \beta(t_1)) = (\alpha(t_2), \beta(t_2))$ . These double points are called *crossings* of the knot. At a given crossing for  $t_1 < t_2$  if  $\gamma(t_1) < \gamma(t_2)$  then it is an *under-crossing* and if  $\gamma(t_1) > \gamma(t_2)$  it is an *over-crossing*. Knots which admit a regular projection are known as *tame knots* (for the precise definition of tame knot see [2]). Thus, for us a knot will always mean a tame-knot.

It is easy to see that each ambient isotopy class of knots in  $S^3$  contains a knot given by a smooth embedding  $\phi$  of  $S^1$  in  $S^3$  which maps the base point  $(0, 1) \in S^1$  to the base point  $(0, 0, 0, 1) \in S^3$  and whose derivative at  $(0, 1)$  is non-zero. Identifying  $\mathbb{R}$  with  $S^1 \setminus \{(0, 1)\}$  and  $\mathbb{R}^3$  with  $S^3 \setminus \{(0, 0, 0, 1)\}$ , one observes that an embedding  $\phi: S^1 \rightarrow S^3$  which is base point preserving and whose derivative at the given base point is non-zero, can be identified with a proper, smooth embedding  $\tilde{\phi}: \mathbb{R} \rightarrow \mathbb{R}^3$  which is monotone (i.e.  $\|\tilde{\phi}(x)\| \rightarrow \infty$  strictly monotonically as  $|x| \rightarrow \infty$ ) outside a closed

interval. Conversely, if  $\psi: \mathbb{R} \rightarrow \mathbb{R}^3$  is a proper, smooth embedding which is monotone outside a closed interval, then  $\psi$  defines an embedding  $\hat{\psi}: S^1 \rightarrow S^3$  (which is smooth everywhere except possibly at infinity) such that  $\hat{\psi}|_{\mathbb{R}} = \psi$  and  $\hat{\psi}((0, 1)) = (0, 0, 0, 1)$ . Two proper, smooth embeddings  $\psi_0$  and  $\psi_1$  of  $\mathbb{R}$  in  $\mathbb{R}^3$  which are monotone outside a closed interval are said to be *equivalent* if  $\hat{\psi}_0$  and  $\hat{\psi}_1$  are ambient isotopic in  $S^3$  via base point preserving diffeomorphisms, which is same as saying that there exists an isotopy  $F: \mathbb{R} \times I \rightarrow \mathbb{R}^3$  between  $\psi_0$  and  $\psi_1$ . A proper, smooth embedding of  $\mathbb{R}$  in  $\mathbb{R}^3$  which is monotone outside a closed interval is also known as a non-compact knot (e.g. see [5]) and it is called a non-compact tame knot if its extension as an embedding of  $S^1$  in  $S^3$  is a tame knot. In this paper, by a knot-type we mean an equivalence class of a non-compact tame knot. Observe that, if  $\psi: \mathbb{R} \rightarrow \mathbb{R}^3$  is an embedding defined as  $\psi(t) = (f(t), g(t), h(t))$ , where  $f(t), g(t)$  and  $h(t)$  are polynomials over  $\mathbb{R}$ , then  $\psi$  is a proper, smooth embedding which is monotone outside a closed interval and it is a non-compact tame knot. If  $K$  denotes the knot-type given by  $\psi$ , then we say that  $\psi$  is a representation (in fact polynomial representation here) of  $K$  or the knot-type  $K$  is represented by the polynomial embedding  $\psi$ . Sometimes, we also say that  $K$  has a polynomial representation by the polynomials  $f(t), g(t)$  and  $h(t)$ .

The question of representing a knot-type by polynomials came up because of Abhyankar's conjecture [1] regarding the existence of non-rectifiable embeddings of affine line  $\mathcal{A}_k^1$  in the affine space  $\mathcal{A}_k^3$ . To bring in the topological point of view, we can assume that our field  $k = \mathbb{C}$ . Recently, Shastri [4] proved that, 'given a knot-type, there exist real polynomials  $f(t), g(t)$  and  $h(t)$  such that the map  $t \mapsto (f(t), g(t), h(t))$  is an embedding of  $\mathbb{C}$  in  $\mathbb{C}^3$  and as an embedding of  $\mathbb{R}$  in  $\mathbb{R}^3$ , it represents the given knot-type'. In other words, every knot-type has a polynomial representation. One may ask now, whether two polynomial representations of a given knot-type are related polynomially or not? (i.e. whether we can define an isotopy between these embeddings through polynomials?). In this paper we have shown that the answer to this question is 'yes'. We have given an alternative proof of the theorem proved in [4]. Our proof involves a suitable  $C^1$ -Weierstrass approximation of all three functions representing the given knot-type inside a closed interval, instead of constructing the third polynomial and then choosing a Weierstrass approximation of first two functions. We have used this method to prove the existence of a polynomial isotopy between two polynomial embeddings which represent the same knot-type.

## 2. Main results

**Theorem 2.1.** *Every knot-type has a polynomial representation.*

*Proof.* Let  $K_\phi$  be the given knot-type, given by a smooth embedding  $\phi: \mathbb{R} \rightarrow \mathbb{R}^3$ , defined as  $\phi(t) = (\alpha(t), \beta(t), \gamma(t))$  such that  $\tilde{\phi} \equiv (\alpha, \beta): \mathbb{R} \rightarrow \mathbb{R}^2$  is a generic immersion, i.e. defines a regular projection of  $K_\phi$ . Up to equivalence we can assume that the derivatives  $\alpha', \beta'$  and  $\gamma'$  of  $\alpha, \beta$  and  $\gamma$  are bounded away from zero outside a closed interval, say  $[a, b]$ . Since we are working with tame knots, there are only finitely many crossings in the image of  $\tilde{\phi}$ , we can choose an interval  $[c, d]$  such that  $\tilde{\phi}([c, d])$  contains all the crossings. Let  $[M_1, M_2] \supset [a, b] \cup [c, d]$  be such that  $\phi([M_1, M_2])$  is contained inside a ball of radius  $R$  with  $\|\phi(M_1)\| = \|\phi(M_2)\| = R$ . Let  $[N_1, N_2]$  be an interval such

that  $\phi([N_1, N_2])$  is contained inside a ball of radius  $2R$  with  $\|\phi(N_1)\| = \|\phi(N_2)\| = 2R$ . By a smooth reparametrization we can assume that  $[M_1, M_2] = [-1/2, 1/2]$  and  $[N_1, N_2] = [-1, 1]$ . Thus  $\phi([-1/2, 1/2])$  and  $\phi([-1, 1])$  are contained inside balls of radius  $R$  and  $2R$  respectively with  $\|\phi(-1/2)\| = \|\phi(1/2)\| = R$  and  $\|\phi(-1)\| = \|\phi(1)\| = 2R$ ; and  $\alpha'$ ,  $\beta'$  and  $\gamma'$  are greater than some positive number (say 1) outside  $[-1/2, 1/2]$  i.e.  $\|\phi\|$  is increasing outside  $[-1/2, 1/2]$ . Consider the restriction of  $\phi$  to  $[-1, 1]$ , i.e.  $\phi|_{[-1, 1]}: [-1, 1] \rightarrow \mathbb{R}^3$ . Since the set of embeddings from a compact, Hausdorff manifold to any manifold forms an open set in the set of all smooth maps with the  $C^1$ -topology (see, [3]), there exists an  $\varepsilon_0 > 0$  such that

$$\psi \in N(\phi, \varepsilon_0) \text{ implies that } \psi \text{ is an embedding of } [-1, 1] \text{ in } \mathbb{R}^3, \quad (*)$$

where

$$N(\phi, \varepsilon_0) = \left\{ \psi: \sup_{t \in [-1, 1]} \{ \|\psi(t) - \phi(t)\|, \|\psi'(t) - \phi'(t)\| \} < \varepsilon_0 \right\}.$$

Let  $\varepsilon < \min\{R/2, \varepsilon_0\}$ . For this  $\varepsilon$ , let  $\psi_1 \equiv (f_1, g_1, h_1)$  be an  $\varepsilon/2$ -Weierstrass  $C^1$  approximation of  $\phi$  inside the interval  $[-1, 1]$ , where  $f_1, g_1$  and  $h_1$  are polynomials. Then  $f'_1, g'_1, h'_1 > 1 - \varepsilon/2$  in  $[-1, -1/2] \cup [1/2, 1]$ . Now, for  $\delta \in (0, \varepsilon/2)$ , we can choose  $N \in \mathbb{Z}^+$  large enough so that  $\psi = \left( f_1 + \frac{\delta}{2N+1} t^{2N+1}, g_1 + \frac{\delta}{2N+1} t^{2N+1}, h_1 + \frac{\delta}{2N+1} t^{2N+1} \right) \equiv (f, g, h)$  is an  $\varepsilon - C^1$  approximation of  $\phi$  inside  $[-1, 1]$  and  $f', g', h' > 0$  outside  $[-1, 1]$ . Note that  $f', g', h' > 1 - \varepsilon$  in  $[-1, -1/2] \cup [1/2, 1]$ . Thus  $\psi' > 0$  outside  $[-1/2, 1/2]$ , i.e.  $\|\psi\|$  is increasing outside  $[-1/2, 1/2]$ .

We shall show that  $\psi: \mathbb{R} \rightarrow \mathbb{R}^3$  is an embedding. Since  $\|\psi'(t)\| \geq \|\phi'(t)\| - \varepsilon$  for  $t \in [-1/2, 1/2]$  and  $\psi' > 0$  for  $t$  outside  $[-1/2, 1/2]$ ,  $\psi$  is an immersion. To show that  $\psi$  is injective we have to show that  $\psi(t_1) \neq \psi(t_2)$  for all  $t_1, t_2 \in \mathbb{R}$ . Now, since  $\psi \in N(\phi, \varepsilon)$ , from (\*) for  $t_1, t_2 \in [-1, 1]$ ,  $\psi(t_1) \neq \psi(t_2)$ . Since outside  $[-1/2, 1/2]$   $\psi' > 0$ , i.e.,  $\|\psi\|$  is increasing, if  $t_1, t_2$  are both in  $(-\infty, -1/2)$  or both in  $(1/2, \infty)$  we have  $\psi(t_1) \neq \psi(t_2)$ . Also, for  $t_1 \in (-\infty, -1)$  and  $t_2 \in (1, \infty)$   $\psi(t_1) \neq \psi(t_2)$ , since  $\psi(t) < 0$  for  $t \in (-\infty, -1)$  and  $\psi(t) > 0$  for  $t \in (1, \infty)$ . Let  $t_1 \in (-\infty, -1)$  and  $t_2 \in [-1/2, 1/2]$ . Then

$$\begin{aligned} \|\psi(t_1) - \psi(t_2)\| &\geq \|\psi(t_1)\| - \|\psi(t_2)\| \\ &\geq (2R - \varepsilon) - (R + \varepsilon) \\ &= R - 2\varepsilon. \end{aligned}$$

Hence,  $\psi(t_1) \neq \psi(t_2)$ . Similarly, we can show that for  $t_1 \in (1, \infty)$  and  $t_2 \in [-1/2, 1/2]$ ,  $\psi(t_1) \neq \psi(t_2)$ . This proves that  $\psi$  is injective. Therefore  $\psi$  is an embedding.

Let  $K_\psi$  denote the knot-type given by  $\psi$ . We will show that  $K_\phi$  and  $K_\psi$  are the same knot-type, i.e., there exists an isotopy between  $\phi$  and  $\psi$ . Define

$$F: \mathbb{R} \times I \rightarrow \mathbb{R}^3$$

as  $F(s, t) = (1 - t)\phi(s) + t\psi(s)$ . Clearly  $F(s, 0) = \phi(s)$  and  $F(s, 1) = \psi(s)$ . Now, for any  $t \in (0, 1)$

$$\begin{aligned} \|(1 - t)\phi + t\psi - \phi\| &= \|-t(\phi - \psi)\| \\ &= t\|\phi - \psi\| \\ &< \varepsilon. \end{aligned}$$

Thus, each  $F_t$  given by  $F_t(s) = F(s, t)$  is an  $\varepsilon - C^1$  approximation of  $\phi$  inside  $[-1, 1]$  and  $F'_t > 0$  outside  $[-1/2, 1/2]$ . By a similar argument as above, we can show that each  $F_t$  is an embedding of  $\mathbb{R}$  in  $\mathbb{R}^3$  and hence  $F$  is an isotopy between  $\phi$  and  $\psi$ . This proves that  $K_\phi$  and  $K_\psi$  are the same knot-type.  $\square$

## DEFINITION 2.2

Two polynomial embeddings  $\phi_0 \equiv (f_0, g_0, h_0)$  and  $\phi_1 \equiv (f_1, g_1, h_1)$  of  $\mathbb{R}$  in  $\mathbb{R}^3$  are said to be polynomially isotopic if there exists an isotopy  $P: \mathbb{R} \times I \rightarrow \mathbb{R}^3$  between  $\phi_0$  and  $\phi_1$  such that  $P_t: \mathbb{R} \rightarrow \mathbb{R}^3$ , given by  $P_t(s) = P(s, t)$  is a polynomial embedding for each  $t \in [0, 1]$ .

**Theorem 2.3.** *Let  $\phi_0 \equiv (f_0, g_0, h_0)$  and  $\phi_1 \equiv (f_1, g_1, h_1)$  be two polynomial embeddings of  $\mathbb{R}$  in  $\mathbb{R}^3$  which represent the same knot-type. Then  $\phi_0$  and  $\phi_1$  are polynomially isotopic.*

To prove this theorem, we observe the following:

*Notation.* For  $\varepsilon \in \mathbb{R}^+$ ,  $N \in \mathbb{Z}^+$  and  $\phi \equiv (f, g, h): \mathbb{R} \rightarrow \mathbb{R}^3$ , let  $\phi_{\varepsilon, N}$  denotes the map  $\phi_{\varepsilon, N}: \mathbb{R} \rightarrow \mathbb{R}^3$  given by  $\phi_{\varepsilon, N}(s) = (f(s), g(s), h(s) + \varepsilon s^{2N+1})$ .

**Lemma 2.4.** *Let  $\phi = (f, g, h): \mathbb{R} \hookrightarrow \mathbb{R}^3$  be a polynomial embedding such that the map  $(f, g): \mathbb{R} \rightarrow \mathbb{R}^2$  is a generic immersion. Then for each positive integer  $N$ , there exists an  $\varepsilon > 0$  such that  $\phi$  and  $\phi_{\varepsilon, N}$  are polynomially isotopic.*

*Proof.* We shall show that the map  $F_t: \mathbb{R} \rightarrow \mathbb{R}^3$  given by

$$F_t(s) = (f(s), g(s), h(s) + t\varepsilon s^{2N+1})$$

is an embedding for each  $t \in [0, 1]$ , so that they define an isotopy between  $\phi$  and  $\phi_{\varepsilon, N}$ . It suffices to show that each  $F_t$  is injective. Consider

$$S = \{(s_1, s_2) \in \mathbb{R}^2 \setminus \Delta \mid (f(s_1), g(s_1)) = (f(s_2), g(s_2))\}.$$

Here,  $\Delta$  denotes the diagonal set in  $\mathbb{R}^2$ . Note that,  $S$  is finite. As  $\phi$  is an embedding,  $h(s_1) \neq h(s_2)$  for  $(s_1, s_2) \in S$ . Choose

$$0 < \varepsilon < \min_{(s_1, s_2) \in S} \left\{ \frac{|h(s_1) - h(s_2)|}{|s_1^{2N+1} - s_2^{2N+1}|} \right\}.$$

Then for this  $\varepsilon$

$$h(s_1) + t\varepsilon s_1^{2N+1} \neq h(s_2) + t\varepsilon s_2^{2N+1}$$

whenever  $(s_1, s_2) \in S$  and  $t \in [0, 1]$ . Thus  $F_t$  is injective for all  $t \in [0, 1]$ . This proves that  $\phi$  is polynomially isotopic to  $\phi_{\varepsilon, N}$ .  $\square$

**Lemma 2.5** *Let  $\phi_0 \equiv (f_0, g_0, h_0)$  be a polynomial embedding of  $\mathbb{R}$  in  $\mathbb{R}^3$ . There exists a polynomial embedding  $\phi_1 \equiv (f_1, g_1, h_1)$  of  $\mathbb{R}$  in  $\mathbb{R}^3$  with  $(f_1, g_1): \mathbb{R} \rightarrow \mathbb{R}^2$  a generic immersion such that  $\phi_0$  and  $\phi_1$  are polynomially isotopic.*

*Proof.* If  $(f_0, g_0): \mathbb{R} \rightarrow \mathbb{R}^2$  is a generic immersion, we have nothing to prove. Now, assume that  $(f_0, g_0): \mathbb{R} \rightarrow \mathbb{R}^2$  is not a generic immersion. Since  $\phi_0$  is a polynomial



embedding, it represents a tame knot. Therefore, we can choose a plane  $P$  such that  $\pi_P \circ \phi$  is a generic immersion (where  $\pi_P: \mathbb{R}^3 \rightarrow P$  is the projection). By linear change of coordinates, we can get a linear automorphism  $A \in GL(3, \mathbb{R})^+$  which sends  $P$  to  $xy$  plane. As  $A \in GL(3, \mathbb{R})^+$  and  $GL(3, \mathbb{R})^+$  is connected,  $A$  is isotopic to  $Id_{\mathbb{R}^3}$  through linear maps. Hence  $\phi_0$  is isotopic to  $\phi_1 = A \circ \phi_0$  through linear maps, i.e., they are polynomially isotopic. From the choice of  $A, (f_1, g_1): \mathbb{R} \rightarrow \mathbb{R}^3$  is a generic immersion.  $\square$

*Proof* (of Theorem 2.3). By Lemmas 2.4 and 2.5, up to polynomial isotopy, we can assume the following:

- (i) Degrees of  $h_0$  and  $h_1$  are odd and sufficiently large so that  $h'_0 > 1$  and  $h'_1 > 1$  outside a suitable interval  $[-M_1, M_1]$  and  $\|\phi_0\|$  and  $\|\phi_1\|$  are increasing outside  $[-M_1, M_1]$ .
- (ii)  $\tilde{\phi}_0: \mathbb{R} \rightarrow \mathbb{R}^2$  and  $\tilde{\phi}_1: \mathbb{R} \rightarrow \mathbb{R}^2$  are generic immersions.

Let  $[a, b]$  be a common interval such that  $\tilde{\phi}_0([a, b])$  and  $\tilde{\phi}_1([a, b])$  contain all the crossings of  $\tilde{\phi}_0$  and  $\tilde{\phi}_1$  respectively. Let  $[N_1, N_2] \supseteq [a, b] \cup [-M_1, M_1]$  be such that  $\phi_0([N_1, N_2])$  and  $\phi_1([N_1, N_2])$  are contained inside a ball of radius  $R$  with  $\|\phi_i(N_1)\| = \|\phi_i(N_2)\| = R$  for  $i = 0, 1$  and  $\|\phi_0\|$  and  $\|\phi_1\|$  are increasing outside  $[N_1, N_2]$ . Without loss of generality we can take  $[N_1, N_2] = [-\frac{1}{2}, \frac{1}{2}]$ , (since we have a linear homeomorphism  $\tau: \mathbb{R} \rightarrow \mathbb{R}$  which maps  $[-\frac{1}{2}, \frac{1}{2}]$  to  $[N_1, N_2]$ , therefore  $\phi_0 \circ \tau$  is polynomially isotopic to  $\phi_1 \circ \tau$  would imply that  $\phi_0$  is polynomially isotopic to  $\phi_1$ ). Let  $\|\phi_i(-1)\| = R + r_i$  and  $\|\phi_i(1)\| = R + r'_i$  for  $i = 0, 1$ . Let  $r = \min_i \{r_i, r'_i\}$ . Now, since  $\phi_0$  and  $\phi_1$  represent the same knot-type, there exists a smooth isotopy  $F: \mathbb{R} \times I \rightarrow \mathbb{R}^3$  such that  $F(s, 0) = \phi_0(s)$  and  $F(s, 1) = \phi_1(s)$  for  $s \in \mathbb{R}$ . Consider  $F|_{[-1, 1] \times I}: [-1, 1] \times I \rightarrow \mathbb{R}^3$ . Since  $I$  is compact, we can choose an  $\varepsilon_0 > 0$  such that  $E_t$  is an embedding of  $[-1, 1]$  in  $\mathbb{R}^3$  whenever  $E_t$  lies inside an  $\varepsilon_0$ -neighbourhood of  $F_t$  (in  $C^1$  topology) for all  $t \in [0, 1]$ . Let  $\varepsilon < \min\{\varepsilon_0, r/2\}$ . For this  $\varepsilon$  let  $P$  be a polynomial map which is an  $\varepsilon$  Weierstrass  $C^1$ -approximation of  $F$  inside the compact set  $[-1, 1] \times I$  such that for every  $t \in I$ ,  $F_t$  is approximated by  $(p_t, q_t, r_t)$  where  $r'_t > 0$  outside  $[-1/2, 1/2]$  (as in the proof of Theorem 2.1, though the proof of Theorem 2.1 is for a fixed  $t$  but easily works for parameters in a compact set). Thus  $P_t: \mathbb{R} \rightarrow \mathbb{R}^3$  is an embedding of  $\mathbb{R}$  in  $\mathbb{R}^3$  for all  $t \in [0, 1]$ . Thus we get a polynomial isotopy between  $P_0$  and  $P_1$ . The proof will be complete, if we produce polynomial isotopies between  $\phi_0$  and  $P_0$  and between  $\phi_1$  and  $P_1$ . Define  $H: \mathbb{R} \times I \rightarrow \mathbb{R}^3$  given by

$$H(s, t) = (1 - t)\phi_0(s) + tP_0(s).$$

Now, from our choice of  $\varepsilon$  and  $P$ , each  $H_t$  defined as  $H_t(s) = (1 - t)\phi_0(s) + tP_0(s)$  is an  $\varepsilon C^1$ -Weierstrass approximation of  $\phi_0$  inside  $[-1, 1]$  and  $\|H_t\|$  is increasing outside  $[-1/2, 1/2]$ . Using a similar argument as in the proof of Theorem 2.1, we can show that each  $H_t$  is an embedding of  $\mathbb{R}$  in  $\mathbb{R}^3$ . Thus,  $H$  defines a polynomial isotopy between  $\phi_0$  and  $P_0$ . Similarly, we can show that  $\phi_1$  and  $P_1$  are polynomially isotopic. This completes the proof of the theorem.  $\square$

### Acknowledgement

The author is thankful to Prof. A R Shastri and Dr A Ranjan for useful discussion. The author would also like to thank the referee for valuable suggestions.

**References**

- [1] Abhyankar S S, On the semigroup of a meromorphic curves in *Int. Sym. Alg. Geom. (part I)* (Kyoto) Kinokuniya, Tokyo (ed) M Nagata (1977) pp. 249–414
- [2] Burde G and Zieschang H, *Knots* (ed.) Walter de Gruyter (1989)
- [3] Hirsh M W, *Differential Topology* (Springer-Verlag) (1975)
- [4] Shastri A R, Polynomial representations of knots, *Tôhoku Math. J.* **44** (1992) 11–17
- [5] Vassiliev V A, Cohomology of knot spaces in *Theory of Singularities and its Applications* (Adv. Sov. Math.) Vol. 1 (1990)

## Row-reduction and invariants of Diophantine equations

N J WILDBERGER

School of Mathematics, University of New South Wales, Sydney 2052 Australia

MS received 30 March 1993; revised 9 February 1994

**Abstract.** To any Diophantine equation with integral coefficients we associate a finitely generated abelian group. The analysis of this group by row-reduction generally leads to simpler equations which are equivalent to the original but often dramatically easier to solve. This method of studying equations is useful over finite fields as well as over  $\mathbb{Q}$ . Some applications and an example are discussed.

**Keywords.** Diophantine equations; row-reduction.

### Introduction

Let  $f$  be a polynomial in the variables  $X_1, \dots, X_n$  with integral coefficients and consider the problem of finding all rational solutions of the equation

$$f(X_1, \dots, X_n) = 0. \quad (1)$$

Historically work on this problem has focused on particular equations of low degree with a few variables. General results that apply to all polynomials or even large classes of polynomials are few and are mostly concerned with the existence of non-zero solutions, where a solution  $(s_1, \dots, s_n)$  of (1) is called non-zero if at least one of the  $s_i$  is non-zero.

For an arbitrary large non-homogeneous polynomial  $f$ , however, it would seem that this basic problem is hopelessly difficult. We hope to show in this paper that this is not so; specifically we will present a method to modify the general equation (1) which often results in a drastic simplification of the equation and sometimes to a complete determination of all solutions. It will be seen that this method can be useful when trying to solve (1) over any field, and in particular over a finite field it is surprisingly powerful considering its simplicity.

Before presenting the method in detail, we make a comment on the relevance of non-zero solutions of (1). Consider the following elementary method of obtaining such solutions. Suppose that  $n \geq 2$ . If one of the variables  $X_i$  occurs in each term of  $f$ , set  $X_i = 0$  and let the other variables be arbitrary with at least one of them non-zero; this is a non-zero solution. Otherwise, pick one of the variables, set it to zero and examine the resulting equation for a variable which occurs in each term. If one occurs and the number of variables still exceeds 1, then we have a non-zero solution as before, otherwise we continue. We eventually get a non-zero solution or arrive at an equation with exactly one variable. If we can find a non-zero solution to this one-variable equation, we have a non-zero solution to the original equation.

We see that if the polynomial  $f(X_1, \dots, X_n)$  has  $k$  terms with  $k < n$ , then (1) has a non-zero solution. The reason is as follows. If after setting a variable to zero the total number of variables has decreased by 2 or more then a non-zero solution may be obtained immediately. Otherwise at each stage of the above algorithm the number of variables exceeds the number of terms, so we can never reach a one-variable equation and must arrive instead at a non-zero solution.

These remarks show that non-zero solutions may often exist for trivial reasons. Define therefore a solution  $(s_1, \dots, s_n)$  of (1) to be non-trivial if all of the  $s_i$  are non-zero. Henceforth in this paper the term solution refers to non-trivial solution.

## 1. The method

Write

$$f(X_1, \dots, X_n) = \sum_{j=1}^k c_j \prod_{i=1}^n X_i^{\alpha_{ij}} \quad (2)$$

where  $c_j$  are non-zero integers and  $\alpha_{ij}$  are non-negative integers. Introduce a new variable  $X_0$ , multiply  $f$  by  $X_0$ , and write  $f(X_0, X_1, \dots, X_n) = X_0 f(X_1, \dots, X_n)$ . To each variable  $X_i$  associate the power vector  $\alpha_i = (\alpha_{i1}, \dots, \alpha_{ik}) \in \mathbf{Z}^k$ . The power vector of  $X_0$  is just  $\alpha_0 = (1, \dots, 1)$ . Let  $A_f \in M((n+1) \times k, \mathbf{Z})$  be the matrix whose rows are  $\alpha_0, \alpha_1, \dots, \alpha_n$  in that order. Call  $A_f$  the power matrix of  $f$ . Let  $G_f$  be the subgroup of the abelian group  $\mathbf{Z}^k$  generated by the vectors  $\alpha_0, \alpha_1, \dots, \alpha_n$ . Call  $G_f$  the power group of  $f$ . Finally let  $c = (c_1, \dots, c_k)$  and call it the coefficient vector of  $f$ .

These definitions are closely connected to the only general change of variable that leaves  $k$  unchanged. Introduce new variables  $Y_0, Y_1, \dots, Y_m$  and suppose that

$$X_j = \prod_{i=0}^m Y_i^{\beta_{ij}} \quad 0 \leq j \leq n \quad (3)$$

for some constants  $\beta_{ij} \in \mathbf{Z}$ . Let  $B \in M((m+1) \times (n+1), \mathbf{Z})$  be defined by  $B = [\beta_{ij}]$ ,  $0 \leq i \leq m, 0 \leq j \leq n$ .

Then

$$\begin{aligned} f(X_0, X_1, \dots, X_n) &= \sum_{j=1}^k c_j \prod_{i=0}^n \left( \prod_{l=0}^m Y_l^{\beta_{li}} \right)^{\alpha_{ij}} \\ &= \sum_{j=1}^k c_j \prod_{l=0}^m Y_l^{\gamma_{lj}} \\ &= h(Y_0, Y_1, \dots, Y_m). \end{aligned} \quad (4)$$

If  $A_h = [\gamma_{lj}] \in M((m+1) \times k, \mathbf{Z})$  is the power matrix of  $h$ , then (4) shows that

$$A_h = B A_f. \quad (5)$$

If  $B$  has a (generalized) left inverse  $B' \in M((n+1) \times (m+1), \mathbf{Z})$ , then the change of variable (3) is invertible and  $A_f = B' A_h$ .

Note that  $h$  has exactly  $k$  terms and the same coefficient vector as  $f$ . Any solution of

$h(Y_0, Y_1, \dots, Y_m) = 0$  immediately gives us via (3) a solution of  $f(X_0, X_1, \dots, X_n) = 0$ . This suggests that we introduce a partial ordering on the set of all polynomials of  $k$  terms with coefficient vector  $c$ . If  $f, h$  are two such polynomials related as above, then we will write  $h < f$ . If  $h < f$  and  $f < h$  then we will say that  $f$  and  $h$  are equivalent and write  $f \sim h$ . If  $G_f$  and  $G_h$  are the power groups of  $f$  and  $h$  respectively, then  $h < f$  if and only if  $G_h \subset G_f$ . Thus  $h \sim f$  if and only if  $G_h = G_f$  and so the equivalence class of a polynomial  $f$  of  $k$  terms is determined completely by its coefficient vector  $c$  and its power group  $G_f$ .

Now any non-trivial subgroup  $G_f$  of  $\mathbf{Z}^k$  must be free abelian of rank  $(r+1) \leq k$  for some non-negative integer  $r$  which we call the rank of  $f$ . If  $r < n$  then some of the variables  $X_i, i > 0$  are redundant in the sense that  $f$  is equivalent to a polynomial  $h$  with strictly fewer variables. This will necessarily occur, for example, if  $k < n$ .

To determine  $G_f$  as precisely as possible, we may row reduce the power matrix  $A_f$ . Since we are dealing with integral matrices, we are allowed to interchange rows, multiply a row by  $-1$ , or add a multiple of one row to another row. We may also permute columns, since this corresponds to relabelling the terms of  $f$ , as long as we remember that the coefficient vector  $c$  also changes thereby.

The result is to obtain a matrix in the row echelon form

$$\begin{vmatrix} 1 & 1 & \cdots & & & 1 \\ 0 & d_1 & * & & \cdots & * \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & & & d_r & * & \cdots * \\ 0 & & \cdots & & & 0 \\ \vdots & & & & & \vdots \\ 0 & & \cdots & & & 0 \end{vmatrix} \quad (6)$$

such that

- 1) row 0 consists entirely of 1's.
- 2)  $1 \leq d_1 \leq \cdots \leq d_r$
- 3) all rows past row  $r$  are zero
- 4) any entry  $s$  above  $d_i$ , but not in row 0 satisfies  $0 \leq s < d_i$
- 5) For any  $1 \leq i \leq r$  and  $j > i$ , the elements of the  $j$ th column from row  $i$  to row  $r$  inclusive generate a subgroup of  $\mathbf{Z}$  which is 0 or whose minimal positive generator  $h$  satisfies  $d_i \leq h$ . In particular any non-zero entry  $s$  lying to the right and possibly below  $d_i$  must have absolute value at least  $d_i$ .

Such a reduced matrix corresponds to an equation equivalent to the original but often considerably simpler. Let us now consider some conditions which ensure that the new equation is easily solvable.

Let  $e_i = (0, \dots, 0, 1, 0, \dots, 0)$  denote the element of  $\mathbf{Z}^k$  with exactly one 1 in the  $i$ th position. If an equation contains a variable  $X_j$  with power vector  $e_i$ , then solving the equation is easy because  $X_j$  appears linearly in only one term. Specifically let all the other variables be arbitrary and non-zero, solve for  $X_j$  and reject cases for which  $X_j$  is zero. This gives us all solutions of the equation.

More generally if an equation has a power group which contains a vector  $e_i$ , then we may make a change of variable to get an equivalent equation which contains a

as a power vector. The above method then gives us all solutions to the modified equation which we can transform back to get all solutions of the original equation.

A related situation occurs when we find that the power group contains a vector  $e$  consisting entirely of 0's and 1's with at least one of each. Then a change of variable as described above results in an equation of the form

$$Y_1 h_1(Y_2, \dots, Y_m) = h_2(Y_2, \dots, Y_m) \quad (7)$$

where both  $h_1$  and  $h_2$  have fewer than  $k$  terms. Let  $Y_2, \dots, Y_m$  be arbitrary and non-zero such that either both  $h_1(Y_2, \dots, Y_m)$  and  $h_2(Y_2, \dots, Y_m)$  are non-zero or both are zero. In the first case we may solve for non-zero  $Y_1$  directly; in the second case  $Y_1$  may be an arbitrary non-zero value. Similar remarks can be made when  $G_f$  contains a vector consisting entirely of 0's and  $d$ 's where the problem reduces to identifying rationals as  $d$ th powers.

As an illustration of the above consider the special case when  $f$  is diagonal; that is of the form

$$f(X_1, \dots, X_k) = c_1 X_1^{\alpha_{11}} + \dots + c_k X_k^{\alpha_{kk}}. \quad (8)$$

The power matrix is

$$A_f = \begin{vmatrix} 1 & 1 & \dots & 1 \\ \alpha_{11} & 0 & \dots & 0 \\ & \alpha_{22} & 0 & \dots & 0 \\ & & \vdots & & \vdots \\ 0 & & & & \alpha_{kk} \end{vmatrix} \quad (9)$$

Suppose one of the  $\alpha_{ii}$ , say  $\alpha_{11}$ , is relatively prime to all the rest. Then by the Chinese Remainder Theorem, we may find integers  $t_1, \dots, t_k, t$  such that  $\alpha_{11}t_1 + 1 = \alpha_{22}t_2 = \dots = \alpha_{kk}t_k = t$ . But then  $e_1 = t\alpha_0 - t_1\alpha_1 - t_2\alpha_2 - \dots - t_k\alpha_k$  so that  $e_1 \in G_f$ . Conversely if  $G_f$  contains a vector  $e_i$  then  $\alpha_{ii}$  is relatively prime to all the other  $\alpha_{jj}$ . The method above then shows how to obtain all rational solutions to  $f(X_1, \dots, X_k) = 0$ . This is the result essentially contained in Wildberger [1]. We may extend this result by noting that  $G_f$  will contain a vector  $e$  consisting entirely of 0's and 1's with at least one of each if and only if the set  $\{\alpha_{ii} | i = 1, \dots, k\}$  can be partitioned into 2 non-empty subsets such that each element of one subset is relatively prime to all the elements of the other subset. In this case the method above shows how to obtain all rational solutions to  $f(X_1, \dots, X_k) = 0$ .

In the above discussion the nature of the coefficient vector  $c$  is largely irrelevant. Equations whose power groups contain vectors like  $e_i$  are insensitive to changes of the coefficients (as long as they all remain non-zero). This motivates us to define a subgroup  $G \subseteq \mathbf{Z}^k$  to be universally solvable if any equation whose power group is  $G$  has at least one solution. It seems of some interest to classify these.

## 2. Congruence equations

Now consider the case of a congruence equation. Let  $p$  be a prime,  $f$  a polynomial as in (2) with  $c_j \not\equiv 0 \pmod{p}$ ,  $j = 1, \dots, k$ , and consider the equation

$$f(X_1, \dots, X_n) \equiv 0 \pmod{p}. \quad (10)$$

A solution  $(s_1, \dots, s_n) \in \mathbf{Z}_p^n$  is non-trivial if and only if  $S_i \not\equiv 0 \pmod{p}$  for all  $i = 1, \dots, n$ . From Fermat's theorem it follows that we need only know the exponents  $\alpha_{ij}$  up to a multiple of  $m = p - 1$ . Thus the power vectors  $\alpha_i$  are in  $\mathbf{Z}_m^k$ , the power matrix  $A_f$  has entries in  $\mathbf{Z}_m$  and the power group  $G_f$  is a subgroup of  $\mathbf{Z}_m^k$ . These objects are images of the corresponding vectors, matrices and groups defined in the rational case under the obvious homomorphisms. There are then only a finite number of equivalence classes of equations since there are only a finite number of subgroups of  $\mathbf{Z}_m^k$ . Row reducing  $A_f$  as before we obtain the description following (6) but now in addition we may arrange that all entries  $s$  satisfy  $0 \leq s < m$  and that each  $d_i$  is a divisor of  $m$ .

Consider for example the case when  $m = 2q$  with  $q$  a prime, and suppose that  $A_f$  is of the form described above. The possible values for  $d_i$  are 1, 2 and  $q$  and we may immediately deduce the following concerning the entries  $s_{ij}$   $0 \leq i \leq n$ ,  $1 \leq j \leq k$  of  $A_f$ . If  $d_i = q$  then  $s_{ij} = 0$  for  $i < j \leq r$  and  $s_{ij} = 0$  or  $q$  for  $r < j$ . If  $d_i = 2$  and  $d_j = q$  and  $s_{ij} = q$  then  $s_{ji} = 0$ . If  $d_i = 1$  and  $d_j = 2$  and  $d_l = q$  then one of  $s_{ij}, s_{il}$  is 0. If  $d_i = q$  for some  $i$ , then the corresponding equation is of the form

$$Y_i^q h_1(Y_1, \dots, \hat{Y}_i, \dots, Y_n) \equiv h_2(Y_1, \dots, \hat{Y}_i, \dots, Y_n) \pmod{p} \quad (11)$$

where  $h_1, h_2$  are polynomials with fewer than  $k$  terms. Since  $Y_i^q \equiv \pm 1 \pmod{p}$ , the original equation is then reduced to two equations with fewer terms. A similar analysis will be possible whenever  $m$  has a small number of prime factors.

### 3. An example

We now illustrate the general procedure with an example. Consider the congruence equation

$$13x^2y + 4x^4z^3 + 5x^3y^5z^6 + 3y^7z^{16} \equiv 0 \pmod{71}. \quad (12)$$

Introduce another variable and rewrite the equation as

$$\begin{aligned} f(X_0, X_1, X_2, X_3) = \\ 13X_0X_1^2X_2 + 4X_0X_1^4X_3^3 + 5X_0X_1^3X_2^5X_3^6 + 3X_0X_2^7X_3^{16} \equiv 0 \pmod{71}. \end{aligned} \quad (13)$$

Then the power matrix of  $f$  is

$$A_f = \begin{vmatrix} 1 & 1 & 1 & 1 \\ 2 & 4 & 3 & 0 \\ 1 & 0 & 5 & 7 \\ 0 & 3 & 6 & 16 \end{vmatrix}, \quad (14)$$

with entries in  $\mathbf{Z}_{70}$ . Row reducing, we find that if

$$B = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 2 & 18 & 32 & -1 \\ -26 & 43 & 10 & -2 \\ 5 & -2 & -1 & 1 \end{vmatrix} \quad (15)$$

then  $BA_f = C$  where

$$C = \begin{vmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 12 \\ 0 & 0 & 0 & 14 \end{vmatrix} \quad (16)$$

It may be checked that  $\det B = 39$ ; since  $(39, 70) = 1$ ,  $B$  is invertible. We introduce new variables  $Y_0, \dots, Y_3$  and set

$$\begin{aligned} X_0 &= Y_0 Y_1^2 Y_2^{-26} Y_3^5 \\ X_1 &= Y_1^{18} Y_2^{43} Y_3^{-2} \\ X_2 &= Y_1^{32} Y_2^{10} Y_3^{-1} \\ X_3 &= Y_1^{-1} Y_2^{-2} Y_3. \end{aligned} \quad (17)$$

Then  $f$  is transformed to

$$13Y_0 + 4Y_0 Y_1 + 5Y_0 Y_2 + 3Y_0 Y_2^{12} Y_3^{14} \equiv 0 \pmod{71} \quad (18)$$

or

$$\begin{aligned} Y_1 &\equiv -4^{-1}(13 + 5Y_2 + 3Y_2^{12} Y_3^{14}) \\ &\equiv 50 + 52Y_2 + 17Y_2^{12} Y_3^{14} \pmod{71}. \end{aligned} \quad (19)$$

Now to go back to the original problem we disregard  $X_0$  and  $Y_0$  and let  $Y_2 = s$ ,  $Y_3 = t$  to get the parametric solution

$$\begin{aligned} x &= (50 + 52s + 17s^{12}t^{14})^{18}s^{43}t^{-2} \\ y &= (50 + 52s + 17s^{12}t^{14})^{32}s^{10}t^{-1} \\ z &= (50 + 52s + 17s^{12}t^{14})^{-1}s^{-2}t \end{aligned} \quad (20)$$

where  $s$  and  $t$  range over all non-zero values in  $\mathbb{Z}_{71}$  that do not satisfy

$$50 + 52s + 17s^{12}t^{14} \equiv 0 \pmod{71}. \quad (21)$$

Note that this equation has fewer terms than the original. We may rewrite this as

$$t^{14} \equiv 43s^{58} + 22s^{59} \pmod{71} \quad (22)$$

which has a solution if and only if

$$(43s^{58} + 22s^{59})^5 \equiv s^{10}(43 + 22s)^5 \equiv 1 \pmod{71}. \quad (23)$$

An easy check shows this last equation has exactly 2 solutions,  $s \equiv 25$  or  $s \equiv 47$ . It thus follows that the original equation has exactly  $70^2 - 14 \cdot 2 = 4872$  solutions and they are all described as above.

This example is not as special as it may look. The reader is encouraged to verify this claim by choosing some equations randomly and utilizing the same procedure to simplify them.



#### 4. Final remarks

We conclude with some general remarks on the range of applicability of the above method, which ultimately depends on row-reduction affecting the power matrix  $A_f$  non-trivially. If the power vectors  $\alpha_i$  are numerous and haphazard, the method works well. If  $n$  is small compared to  $k$  then row-reduction will accomplish little (the extreme case here being a single variable equation with many terms.) However in this case the possibility arises of a useful preliminary linear change of variable which will decrease  $k$ . We have in this paper ignored linear changes of variable, but a complete theory ought to take into account both linear and multiplicative changes of variable together with their interactions. The method is also ineffective when the power vectors  $\alpha_i$  do not interact with each other or with  $\alpha_0$ . This will happen typically in the diagonal case when the exponents have a large number of common divisors. The worst case is when all exponents are identical, so the equation may be said to be of Fermat type. From our point of view then equations of Fermat type are not only historical curiosities but are the purest examples of difficult Diophantine equations.

#### Reference

- [1] Wildberger N J, A soluble Diophantine equation, *Math-Mag.* **49** (1976) pp. 200–201



## Positive values of non-homogeneous indefinite quadratic forms of type (1,4)

V C DUMIR and RANJEET SEHMI\*

Centre for Advanced Studies on Mathematics, Panjab University, Chandigarh 160 014, India

\*Department of Applied Sciences, Punjab Engineering College, Chandigarh 160 012, India

MS received 18 September 1993

**Abstract.** Let  $\Gamma_{r,n-r}$  denote the infimum of all numbers  $\Gamma > 0$  such that for any real indefinite quadratic  $Q$  in  $n$  variables of type  $(r, n-r)$ , determinant  $D \neq 0$  and real numbers  $c_1, \dots, c_n$  there exist  $(x_1, \dots, x_n) \equiv (c_1, \dots, c_n) \pmod{1}$  satisfying

$$0 < Q(x_1, \dots, x_n) \leq (\Gamma|D|)^{1/n}.$$

All the values of  $\Gamma_{r,s}$  are known except  $\Gamma_{1,4}$ . It is shown that

$$8 \leq \Gamma_{1,4} \leq 16.$$

**Keywords.** Quadratic forms; Birch reduction; equivalent forms.

### 1. Introduction

Let  $Q(x_1, \dots, x_n)$  be a real indefinite quadratic form in  $n$  variables of type  $(r, n-r)$  and determinant  $D \neq 0$ . Blaney [9] has shown that there exist  $\Gamma$ , independent of  $Q$  and depending only on  $n$  and  $r$  such that given any real numbers  $C_n$  there exist  $(x_1, \dots, x_n) \equiv (c_1, \dots, c_n) \pmod{1}$  such that

$$0 < Q(x_1, \dots, x_n) \leq (\Gamma|D|)^{1/n}. \quad (1)$$

Let  $\Gamma_{r,n-r}$  denote the infimum of all such numbers  $\Gamma$ . In this notation the following results are known:

$\Gamma_{1,1} = 4$ , Davenport-Heilbronn [11]

$\Gamma_{2,1} = 4$ , Blaney [10] and Barnes [7]

$\Gamma_{1,2} = 8$ ,  $\Gamma_{3,1} = \frac{16}{3}$ ,  $\Gamma_{2,2} = 16$ , Dumir [12], [13], [14]

$\Gamma_{1,3} = 16$ , Dumir and Hans-Gill [15]

$\Gamma_{3,2} = 16$ ,  $\Gamma_{4,1} = 8$ , Hans-Gill and Madhu Raka [17], [18]

$\Gamma_{r,n-r}$  for  $s = 2r - n = 0, \pm 1, 2, 3$ , Bambah, Dumir and Hans-Gill [4], [5], [6]

$\Gamma_{r,r+2}$  and  $\Gamma_{r,r+3}$  for  $r \geq 3$ , Aggarwal and Gupta [1], [2]

$\Gamma_{r+4,r}$  for  $r \geq 1$ , Aggarwal and Gupta [3]

$\Gamma_{2,5}$ , Dumir and Sehmi [16].

\*This paper forms a part of her Ph.D. dissertation accepted by Panjab University, Chandigarh written under the supervision of V C Dumir

Dumir, Hans-Gill and Woods in a paper to appear in *J. Number Theory* have proved that  $\Gamma_{r,n-r}$  depends only on signature  $s = 2r - n \pmod{8}$  for  $n \geq 6$ . Thus, the value of  $\Gamma_{r,n-r}$  is known except for  $\Gamma_{2,4}$  and  $\Gamma_{1,4}$ .  $\Gamma_{2,4} = \frac{64}{3}$  (unpublished) has been proved by Dumir, Hans-Gill, Sehmi and Woods. The exact value of  $\Gamma_{1,4}$  is still unknown. From the work of Jackson [19] it follows that  $\Gamma_{1,4} \leq 32$ . It is conjectured that  $\Gamma_{1,4} = 8$ .

If  $Q = x_1x_2 - (x_3^2 + x_4^2 + x_5^2 + x_3x_4 + x_3x_5 + x_4x_5)$  or  $x_1x_2 - \frac{1}{4}(x_2^2 + x_3^2 + x_4^2 + x_5^2)$  and  $(c_1, \dots, c_5) = (0, \dots, 0)$  or  $(\frac{1}{2}, \dots, \frac{1}{2})$  respectively, then the inequality (1) is not soluble for  $\Gamma < 8$  and is soluble for  $\Gamma = 8$  so that  $\Gamma_{1,4} \geq 8$ . In this paper we show that  $\Gamma_{1,4} \leq 16$ . More precisely we prove

**Theorem.** Let  $Q(x_1, \dots, x_5)$  be a real indefinite quadratic form of type (1, 4) and determinant  $D \neq 0$ . Then given any real numbers  $c_1, \dots, c_5$ , there exist  $(x_1, \dots, x_5) \equiv (c_1, \dots, c_5) \pmod{1}$  such that

$$0 < Q(x_1, \dots, x_5) < (16|D|)^{1/5}.$$

## 2. Some Lemmas

In the course of the proof we shall use the following Lemmas:

**Lemma 1.** Let  $\phi(x, y, z)$  be a positive definite quadratic form of determinant  $D \neq 0$ . Then there exist integers  $x, y, z$  such that

$$0 < \phi(x, y, z) \leq (2|D|)^{1/3},$$

with equality if and only if  $\phi \sim \rho(x^2 + y^2 + z^2 + xy + yz + zx)$ ,  $\rho > 0$ .

This is a result of Gauss and Seeber.

**Lemma 2.** If  $\psi(x, y)$  is a positive definite binary quadratic form of determinant  $\delta$ , then

$$\psi(x, y) \sim Ax^2 + Bxy + Cy^2,$$

where

$$0 \leq B \leq A \leq C \text{ and } AC \leq \frac{4}{3}\delta.$$

**Lemma 3.** Let  $\alpha, \beta, \gamma$  be real numbers with  $\gamma > 1$ . Let  $m$  be the integer defined by  $m < \gamma \leq m + 1$ . Then for any real  $x_0$ ,

(i) there exists  $x \equiv x_0 \pmod{1}$  such that

$$0 < (x + \alpha)^2 + \beta < \gamma,$$

provided

$$\frac{-m^2}{4} < \beta < \gamma - \frac{1}{4}.$$

(ii) There exists  $x \equiv x_0 \pmod{1}$  such that

$$0 < -(x + \alpha)^2 + \beta < \gamma,$$

provided

$$\frac{1}{4} < \beta < \frac{m^2}{4} + \gamma.$$

This is a result of Dumir ([12], [13]).

We shall use the following conventions:

We say that the inequality

$$\alpha < Q(x_1, \dots, x_n) \leq \beta \quad (2.1)$$

is soluble if for any given real numbers  $c_1, \dots, c_n$ , there exist  $(x_1, \dots, x_n) \equiv (c_1, \dots, c_n) \pmod{1}$  satisfying (2.1). If there are integers  $u_1, \dots, u_n$  such that  $Q(u_1, \dots, u_n) = \gamma \neq 0$ , then we say that  $Q$  represents  $\gamma$ .

**Lemma 4.** Let  $Q(x_1, \dots, x_n)$  be a zero form of determinant  $D \neq 0$ . Let  $\alpha, \beta$  be real numbers satisfying

$$\beta - \alpha > 2|D|^{1/n}.$$

Then

$$\alpha < Q(x_1, \dots, x_n) < \beta$$

is soluble.

This is a result of Jackson [19].

**Lemma 5.** Let  $\alpha, \beta, \gamma$  be real numbers with  $\alpha > 0, \gamma > 0$ . Let  $2h, k$  be integers such that

$$|h - k^2\alpha| + \frac{1}{2} < \gamma.$$

Suppose that either  $\alpha \neq h/k^2$  or  $\beta \not\equiv h/k \pmod{1/k, 2\alpha}$ , i.e.,  $\beta - h/k$  is not an integral linear combination of  $1/k$  and  $2\alpha$ . Then for any real number  $v$ , there exist integers  $x, y$  satisfying

$$0 < \pm x + \beta y \pm \alpha y^2 + v < \gamma. \quad (2.2)$$

This is a result of Macbeath [20].

### 3. Reduction

If  $Q$  is an incommensurable quadratic form, then the result follows from results of Margulis [21] and Watson [22]. So we can suppose that  $Q$  is a rational form of determinant  $D \neq 0$ . By Meyer's Theorem, it is a zero form. Using Birch reduction [8] and homogeneity we can suppose that

$$Q = (x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5)x_2 - \phi(x_3, x_4, x_5),$$

where  $\phi$  is a positive definite ternary quadratic form with determinant  $\Delta = 4|D|$ . Let  $d^5 = 16|D|$ .

**Lemma 6.** If  $\phi(x_3, x_4, x_5)$  represents  $a$  with  $0 < a < d/3$  or  $d/2.68 \leq a < d/2$ , the

$$0 < (x_1 + \dots)x_2 - \phi(x_3, x_4, x_5) < d, \quad (3.1)$$

is soluble.

*Proof.* Without loss of generality we can suppose that the representation of  $a$  by  $\phi$  is primitive. Replacing  $\phi$  by an equivalent form we can suppose that  $\phi(1, 0, 0) = a$  and write

$$\phi(x_3, x_4, x_5) = a(x_3 + h_4x_4 + h_5x_5)^2 + \psi(x_4, x_5),$$

where  $\psi$  is a positive definite form with determinant  $\delta = \Delta/a$ . Now (3.1) can be written as

$$0 < -a(x_3 + \dots)^2 + (x_1 + a'_2x_2 + a'_4x_4 + a'_5x_5)x_2 - \psi(x_4, x_5) < d \quad (3.2)$$

Let  $m$  be the integer defined by  $m < d/a \leq m+1$ , then  $m \geq 3$ . By Lemma 3, (3.2) is soluble if we can solve

$$a/4 < (x_1 + a'_2x_2 + \dots)x_2 - \psi(x_4, x_5) < d + \frac{1}{4}m^2a. \quad (3.3)$$

Since  $(x_1 + \dots)x_2 - \psi(x_4, x_5)$  is a zero form with determinant  $-D/a$ , (3.3) is soluble by Lemma 4, if

$$d + \frac{1}{4}(m^2 - 1)a > 2(|D|/a)^{1/4} = (d^5/a)^{1/4},$$

which is satisfied for  $m \geq 3$  and also if  $m = 2$  and  $2 < d/a \leq 2.68$ . This proves the Lemma.

*Remark 1.* By Lemma 6, we can suppose that if  $\phi$  represents  $a > 0$ , then  $a \geq \frac{d}{2}$  or

$$\frac{d}{3} \leq a < \frac{d}{2.68}.$$

$$\text{Let } a = \min \{ \phi(X) : 0 \neq X \in \mathbb{Z}^3 \}. \quad (3.4)$$

By Lemma 1,  $\phi$  represents  $a$  primitively with

$$0 < a \leq (2\Delta)^{1/3} = (8|D|)^{1/3} = (d^5/2)^{1/3}. \quad (3.5)$$

By a unimodular transformation we can suppose that  $\phi(1, 0, 0) = a$  and write

$$\phi(x_3, x_4, x_5) = a(x_3 + h_4x_4 + h_5x_5)^2 + \psi(x_4, x_5),$$

where  $\psi$  is a positive definite binary form with determinant  $\delta = \Delta/a = 4|D|/a$ . By Lemma 2, we can suppose that

$$\psi(x_4, x_5) = Ax_4^2 + Bx_4x_5 + Cx_5^2 = A(x_4 + \lambda x_5)^2 + tx_5^2, \quad (3.6)$$

where

$$0 \leq B \leq A \leq C \text{ and } 0 < AC \leq \frac{4\delta}{3} = 16|D|/3a = d^5/3a, \quad (3.7)$$

Further, we can suppose that  $-\frac{1}{2} < h_4, h_5 \leq \frac{1}{2}, -\frac{1}{2} < a_i \leq \frac{1}{2}$  for each  $i$ . Since  $A + h_4^2 a$  is a value of  $\phi$  and  $a$  is the minimum value, we have

$$a \leq A + h_4^2 a \leq A + \frac{a}{4},$$

so that

$$A \geq \frac{3a}{4} \text{ and if } h_4 = 0 \text{ then } A \geq a. \quad (3.9)$$

We need to show that

$$0 < (x_1 + a_2 x_2 + \dots) x_2 - a(x_3 + \dots)^2 - A x_4^2 - B x_4 x_5 - C x_5^2 < d, \quad (3.10)$$

is soluble. Without loss of generality we can suppose that  $-\frac{1}{2} < c_i \leq \frac{1}{2}$  for each  $i$ .

#### 4. Proof of the Theorem

*Lemma 7. The inequality (3.10) is soluble except when*

- (i)  $c_2 \not\equiv 0 \pmod{1}$  and  $d \leq \frac{1}{2}$ , or
- (ii)  $c_2 \equiv 0 \pmod{1}$ ,  $d \leq 1$ ,  $a < \frac{1}{2}$  and  $a + d \leq 1$  or
- (iii)  $c_2 \equiv 0 \pmod{1}$ ,  $d \leq 1$ ,  $a = \frac{1}{2}$ ,  $h_4 = h_5 = 0$  and  $(a_3, c_3) = (0, \frac{1}{2})$  or  $(\frac{1}{2}, 0)$ .

*Proof.* Choose  $(x_3, x_4, x_5) = (c_3, c_4, c_5)$ ,  $x_2 = \pm 1$  if  $c_2 \equiv 0 \pmod{1}$   $0 < |x_2| \leq \frac{1}{2}$  if  $c_2 \not\equiv 0 \pmod{1}$ . Then choose  $x_1 \equiv c_1 \pmod{1}$  in such a way that

$$0 < (x_1 + \dots) x_2 - a(x_3 + \dots)^2 - A x_4^2 - B x_4 x_5 - C x_5^2 \leq |x_2|,$$

so that (3.10) is soluble if  $c_2 \equiv 0 \pmod{1}$  and  $d > 1$  or if  $c_2 \not\equiv 0 \pmod{1}$  and  $d > \frac{1}{2}$ .

Now suppose that  $c_2 \equiv 0 \pmod{1}$  and  $d \leq 1$ . Choose  $(x_4, x_5) \equiv (c_4, c_5) \pmod{1}$  arbitrarily and  $x_2 = \pm 1$ . Write  $x_1 = x + c_1$  and  $x_3 = y + c_3$ . Then (3.10) is soluble if we can find integers  $x$  and  $y$  such that

$$0 < \pm x + \beta y - a y^2 + v < d, \quad (4.1)$$

where  $\beta = \pm a_3 - 2a(c_3 + h_4 x_4 + h_5 x_5)$  and  $v$  is some constant. Since

$$|\frac{1}{2} - a| + \frac{1}{2} < d$$

is satisfied for  $a \geq \frac{1}{2}$  and for  $a < \frac{1}{2}$  unless  $a + d \leq 1$ , therefore taking  $h = \frac{1}{2}$ ,  $k = 1$ ,  $\alpha = a$  and  $\gamma = d$ , (4.1) is soluble by Lemma 5 unless  $a < \frac{1}{2}$  and  $a + d \leq 1$  or  $a = \frac{1}{2}$  and

$$\pm a_3 - 2a(c_3 + h_4 x_4 + h_5 x_5) \equiv \frac{1}{2} \pmod{1}. \quad (4.2)$$

Since  $-\frac{1}{2} < h_4 \leq \frac{1}{2}$ , taking  $x_4 = c_4$  and  $1 + c_4$ , we get  $h_4 = 0$ . Similarly  $h_5 = 0$ . For  $h_4 = h_5 = 0$ , (4.2) implies that  $(a_3, c_3) = (0, \frac{1}{2})$  or  $(\frac{1}{2}, 0)$ . This proves the lemma.

*Remark 2.* If  $a \geq \frac{d}{2}$ , then (3.5) implies that  $d > \frac{1}{2}$  or  $d = 2a = 2(d^5/2)^{1/3} = \frac{1}{2}$ , in which

case by Lemma 1,

$$Q = (x_1 + \dots)x_2 - \frac{1}{4}(x_3 + \frac{1}{2}x_4 + \frac{1}{2}x_5)^2 - \frac{3}{16}x_4^2 - \frac{1}{8}x_4x_5 - \frac{3}{16}x_5^2.$$

Thus by Lemma 7 and Remark 1, we are left with the following cases:

(i)  $c_2 \not\equiv 0 \pmod{1}$ ,  $d \leq \frac{1}{2}$  and  $\frac{d}{3} \leq a < \frac{d}{2 \cdot 68}$ .

(ii)  $c_2 \not\equiv 0 \pmod{1}$  and

$$Q = (x_1 + \dots)x_2 - \frac{1}{4}(x_3 + \frac{1}{2}x_4 + \frac{1}{2}x_5)^2 - \frac{3}{16}x_4^2 - \frac{1}{8}x_4x_5 - \frac{3}{16}x_5^2$$

(iii)  $c_2 \equiv 0 \pmod{1}$ ,  $d \leq 1$ ,  $a = \frac{1}{2}$ ,  $h_4 = h_5 = 0$ ,  $(a_3, c_3) = (0, \frac{1}{2})$  or  $(\frac{1}{2}, 0)$ .

(iv)  $c_2 \equiv 0 \pmod{1}$ ,  $d \leq 1$ ,  $a < \frac{1}{2}$ ,  $a + d \leq 1$ .

Let  $m$  be the integer satisfying  $m < d/a \leq m + 1$ , then  $m = 1$  or  $2$  and

$$\frac{d}{m+1} \leq a \leq \left(\frac{d^5}{2}\right)^{1/3} \text{ so that } d^2 \geq \frac{2}{(m+1)^3}. \quad (4.3)$$

The inequality (3.10) can be written as

$$0 < -a(x_3 + \dots)^2 + (x_1 + a'_2x_2 + a'_4x_4 + a'_5x_5)x_2 - Ax_4^2 - Bx_4x_5 - Cx_5^2 < d. \quad (4.4)$$

By Lemma 3, it is soluble if we can solve

$$\frac{a}{4} < (x_1 + \dots)x_2 - Ax_4^2 - Bx_4x_5 - Cx_5^2 < d + \frac{1}{4}m^2a. \quad (4.5)$$

It can be written as

$$0 < -A(x_4 + \dots)^2 + (x_1 + a''_2x_2 + a''_5x_5)x_2 - tx_5^2 - \frac{a}{4} < d + \frac{1}{4}(m^2 - 1)a, \quad (4.6)$$

where  $t$  is as in (3.8).

Let  $K$  be the integer satisfying  $K < \frac{d + \frac{1}{4}(m^2 - 1)a}{A} \leq K + 1$  then it is easy to see that

$$K \geq \begin{cases} 2 & \text{if } d \leq \frac{1}{2}, \\ 1 & \text{if } d \leq 1. \end{cases} \quad (4.7)$$

By Lemma 3, (4.6) is soluble if we can solve

$$A/4 < (x_1 + \dots)x_2 - tx_5^2 - \frac{a}{4} < d + \frac{1}{4}(m^2 - 1)a + \frac{1}{4}K^2A,$$

or if we can solve

$$0 < (x_1 + \dots)x_2 - tx_5^2 - \frac{1}{4}(a + A) < d + \frac{1}{4}(m^2 - 1)a + \frac{1}{4}(K^2 - 1)A. \quad (4.8)$$



**Lemma 8.** For  $m=2$  and  $K \geq 2$ , (4.8) is soluble. In particular if  $c_2 \not\equiv 0 \pmod{1}$ ,  $d \leq \frac{1}{2}$  and  $\frac{d}{3} \leq a < \frac{d}{2.68}$ , then (4.8) is soluble.

*Proof.* From  $m=2$ , (4.8) is soluble by Lemma 4, if

$$d + \frac{3}{4}a + \frac{1}{4}(K^2 - 1)A > 2\left(\frac{t}{4}\right)^{1/3} = \left(\frac{d^5}{2aA}\right)^{1/3}, \quad (\text{by (3.8)})$$

or if

$$\left[d + \frac{3a}{4} + \frac{1}{4}(K^2 - 1)A\right]^3 aA > \frac{d^5}{2},$$

which is satisfied for  $K \geq 2$ , since

$$\begin{aligned} \left[d + \frac{3a}{4} + \frac{1}{4}(K^2 - 1)A\right]^3 aA &\geq \left[\left(d + \frac{3a}{4}\right)\left(\frac{K+3}{4}\right)\right]^3 \left(\frac{a}{K+1}\right)\left(d + \frac{3a}{4}\right) \\ &\geq \left(\frac{5d}{4}\right)^4 \frac{d}{3} \left(\frac{K+3}{4}\right)^3 \frac{1}{K+1} \geq \left(\frac{5}{4}\right)^7 \frac{d^5}{9} > \frac{d^5}{2}. \end{aligned}$$

If  $c_2 \not\equiv 0 \pmod{1}$ ,  $d \leq \frac{1}{2}$  and  $\frac{d}{3} \leq a \leq \frac{d}{2.68}$ , then  $m=2$  and by (4.7),  $K \geq 2$ , therefore (4.8) is soluble in this case.

This completes the lemma.

**Lemma 9.** For  $c_2 \not\equiv 0 \pmod{1}$  and

$$Q = (x_1 + \dots)x_2 - \frac{1}{4}(x_3 + \frac{1}{2}x_4 + \frac{1}{2}x_5)^2 - \frac{3}{16}x_4^2 - \frac{1}{8}x_4x_5 - \frac{3}{16}x_5^2,$$

(4.8) is soluble.

*Proof.* In this case  $d = \frac{1}{2}$ ,  $a = \frac{1}{4}$ ,  $A = \frac{3}{16}$ , so that  $m=1$ ,  $K=2$  and  $d + \frac{1}{4}(m^2 - 1)a + \frac{1}{4}(K^2 - 1)A = d + \frac{3A}{4} = \frac{1}{2} + \frac{9}{64} > \frac{1}{2}$ . Therefore proceeding as in Lemma 7, it is easy to see that (4.8) is soluble.

In view of Lemma 8, Lemma 9 and Remark 2, the case  $c_2 \not\equiv 0 \pmod{1}$  is finished.

**5.**  $c_2 \equiv 0 \pmod{1}$ ,  $d \leq 1$ ,  $a = \frac{1}{2}$ ,  $h_4 = h_5 = 0$ ,  $(a_3, c_3) = (0, \frac{1}{2})$  or  $(\frac{1}{2}, 0)$

**Lemma 10.** For  $a = \frac{1}{2}$ ,  $h_4 = h_5 = 0$ ,  $(a_3, c_3) = (0, \frac{1}{2})$  or  $(\frac{1}{2}, 0)$  (3.10) is soluble unless

- (i)  $A = C = \frac{1}{2}$ ,  $B = 0$ ,  $(a_4, c_4) = (0, \frac{1}{2})$  or  $(\frac{1}{2}, 0)$  and  $(a_5, c_5) = (0, \frac{1}{2})$  or  $(\frac{1}{2}, 0)$  or
- (ii)  $A = \frac{1}{2}$ ,  $B = 0$ ,  $C = 1$ ,  $\pm a_5 - 2c_5 \equiv 0 \pmod{1}$  and  $(a_4, c_4) = (0, \frac{1}{2})$  or  $(\frac{1}{2}, 0)$ .

*Proof.* (3.10) can be written as

$$0 < (x_1 + \dots)x_2 - \frac{1}{2}x_3^2 - Ax_4^2 - Bx_4x_5 - Cx_5^2 < d. \quad (5.1)$$

By (3.7) and (3.9)

$$\frac{1}{2} = a \leq A \leq \left(\frac{d^5}{3a}\right)^{1/2} = \left(\frac{2d^5}{3}\right)^{1/2}. \quad (5.2)$$

Choose  $(x_3, x_5) \equiv (c_3, c_5) \pmod{1}$  arbitrarily,  $x_2 = \pm 1$ . Write  $x_1 = x + c_1$ ,  $x_4 = y + c_4$ , then (5.1) reduces to an inequality of the type (2.2) with  $\alpha = A$  and  $\beta = \pm a_4 - 2Ac_4 - Bx_5$ . Since by (5.2)

$$|\frac{1}{2} - A| + \frac{1}{2} = A \leq \left(\frac{2d^5}{3}\right)^{1/2} < d \quad \text{for } d \leq 1,$$

taking  $h = \frac{1}{2}$ ,  $k = 1$ , it follows from Lemma 5 that (5.1) is soluble unless  $A = \frac{1}{2}$  and  $\beta \equiv h/k \pmod{(1/k, 2\alpha)}$  or

$$\pm a_4 - c_4 - Bx_5 \equiv \frac{1}{2} \pmod{1}. \quad (5.3)$$

Taking  $x_5 = c_5$  and  $1 + c_5$  we get  $B \equiv 0 \pmod{1}$ . Since  $0 \leq B \leq A = \frac{1}{2}$ , we have  $B = 0$ . For  $B = 0$ , (5.3) gives  $(a_4, c_4) = (0, \frac{1}{2})$  or  $(\frac{1}{2}, 0)$ . For  $A = \frac{1}{2}$  and  $B = 0$ , by (3.7), we have

$$\frac{1}{4} = A^2 \leq AC = \frac{C}{2} = \det \psi = 8|D| = \frac{d^5}{2},$$

or

$$\frac{1}{2} \leq C = d^5. \quad (5.4)$$

Choose  $(x_3, x_4) = (c_3, c_4) \pmod{1}$  arbitrarily,  $x_2 = \pm 1$ . Write  $x_1 = x + c_1$ ,  $x_5 = y + c_5$ , then (5.1) reduces to an inequality of the type (2.2) with  $\alpha = C$ ,  $\gamma = d$  and  $\beta = \pm a_5 - 2c_5$ . Take  $k = 1$  and  $h = \frac{1}{2}$  if  $C = d^5 < d$  and  $h = 1$  if  $C = d^5 = d$ . Then it follows from Lemma 5 that (5.1) is soluble unless  $C = \frac{1}{2}$  or 1 and  $\beta \equiv h/k \pmod{(1/k, 2\alpha)}$ . That is, (5.1) is soluble unless (i)  $C = \frac{1}{2}$  and  $(a_5, c_5) = (0, \frac{1}{2})$  or  $(\frac{1}{2}, 0)$  or (ii)  $C = 1$  and  $\pm a_5 - 2c_5 \equiv 0 \pmod{1}$ . This proves the lemma.

*Lemma 11.* For  $a = A = C = \frac{1}{2}$ ,  $B = h_4 = h_5 = 0$ ,  $(a_3, c_3) = (0, \frac{1}{2})$  or  $(\frac{1}{2}, 0)$ ,  $(a_4, c_4) = (0, \frac{1}{2})$  or  $(\frac{1}{2}, 0)$  and  $(a_5, c_5) = (0, \frac{1}{2})$  or  $(\frac{1}{2}, 0)$ , (3.10) is soluble.

*Proof.* Here

$$Q = (x_1 + \dots)x_2 - \frac{1}{2}(x_3^2 + x_4^2 + x_5^2) \quad (5.5)$$

so that

$$d^5 = 16|D| = \frac{1}{2} \text{ or } d = 0.8705\dots$$

We need to consider the following cases:

- (i)  $c_3 = c_4 = c_5 = \frac{1}{2}$  and  $a_3 = a_4 = a_5 = 0$ .
- (ii)  $c_3 = 0$ ,  $c_4 = c_5 = \frac{1}{2}$  and  $a_3 = \frac{1}{2}$ ,  $a_4 = a_5 = 0$ .
- (iii)  $c_3 = c_4 = 0$ ,  $c_5 = \frac{1}{2}$  and  $a_3 = a_4 = \frac{1}{2}$ ,  $a_5 = 0$ .
- (iv)  $c_3 = c_4 = c_5 = 0$  and  $a_3 = a_4 = a_5 = \frac{1}{2}$ .

Let  $f_n = n|c_1| + n^2 a_2$  and  $g_n = -n|c_1| + n^2 a_2$ . In each case depending on the values of  $c_1$  and  $a_2$ , (3.10) can be satisfied by choosing suitable values of  $x_1, \dots, x_5$  as given

Case (i)  $c_3 = c_4 = c_5 = \frac{1}{2}$  and  $a_3 = a_4 = a_5 = 0$ .

$$Q = (x_1 + a_2 x_2) x_2 - \frac{1}{2}(x_3^2 + x_4^2 + x_5^2)$$

|                            | Range   | $x_1$         | $x_1 x_2$    | $x_3$         | $x_4$         | $x_5$         | $Q$                  |
|----------------------------|---|---------------|--------------|---------------|---------------|---------------|----------------------|
|                            | $f_1 > \frac{3}{8}$                               | $c_1$         | $ c_1 $      | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $f_1 - \frac{3}{8}$  |
| $f_1 \leq \frac{3}{8}$     | $-\frac{5}{8} < g_1 < d - \frac{5}{8}$            | $\pm 1 + c_1$ | $1 -  c_1 $  | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $g_1 + \frac{5}{8}$  |
| $f_1 \leq \frac{3}{8}$     | $g_2 < d + \frac{3}{8}$                           | $c_1$         | $-2 c_1 $    | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $g_2 - \frac{3}{8}$  |
|                            | $g_2 \leq d + \frac{3}{8}, f_2 > \frac{11}{8}$    | $c_1$         | $2 c_1 $     | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $f_2 - \frac{11}{8}$ |
| $g_1 \geq d - \frac{5}{8}$ | $g_2 \geq d + \frac{3}{8}, f_2 \leq \frac{11}{8}$ | $c_1$         | $-3 c_1 $    | $\frac{3}{2}$ | $\frac{3}{2}$ | $\frac{1}{2}$ | $g_3 - \frac{19}{8}$ |
| $f_1 \leq \frac{3}{8}$     | $f_1 < d - \frac{5}{8}$                           | $\pm 1 + c_1$ | $1 +  c_1 $  | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $f_1 + \frac{5}{8}$  |
|                            | $f_1 \geq d - \frac{5}{8} > f_2$                  | $\pm 1 + c_1$ | $2 + 2 c_1 $ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{3}{2}$ | $f_2 + \frac{5}{8}$  |
| $g_1 \leq -\frac{5}{8}$    | $f_1, f_2 \geq d - \frac{5}{8} > f_3$             | $\pm 1 + c_1$ | $3 + 3 c_1 $ | $\frac{3}{2}$ | $\frac{3}{2}$ | $\frac{1}{2}$ | $f_3 + \frac{5}{8}$  |
|                            | $f_1, f_2, f_3 \geq d - \frac{5}{8}$              | $\pm 1 + c_1$ | $2 - 2 c_1 $ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $g_2 + \frac{13}{8}$ |

Case (ii)  $c_3 = 0, c_4 = c_5 = \frac{1}{2}$  and  $a_3 = \frac{1}{2}, a_4 = a_5 = 0$ .

$$Q = (x_1 + a_2 x_2 + \frac{1}{2} x_3) x_2 - \frac{1}{2}(x_3^2 + x_4^2 + x_5^2).$$

|                        | Range   | $x_1$         | $x_1 x_2$    | $x_3$ | $x_4$         | $x_5$         | $Q$                 |
|------------------------|---|---------------|--------------|-------|---------------|---------------|---------------------|
|                        | $f_1 > \frac{1}{4}$                               | $c_1$         | $ c_1 $      | 0     | $\frac{1}{2}$ | $\frac{1}{2}$ | $f_1 - \frac{1}{4}$ |
| $f_1 \leq \frac{1}{4}$ | $-\frac{3}{4} < g_1 < d - \frac{3}{4}$            | $\pm 1 + c_1$ | $1 -  c_1 $  | 0     | $\frac{1}{2}$ | $\frac{1}{2}$ | $g_1 + \frac{3}{4}$ |
|                        | $g_1 \geq d - \frac{3}{4}$                        | $c_1$         | $-2 c_1 $    | 0     | $\frac{1}{2}$ | $\frac{1}{2}$ | $g_2 - \frac{1}{4}$ |
|                        | $g_1 \leq -\frac{3}{4}, f_1 < d - \frac{3}{4}$    | $\pm 1 + c_1$ | $1 +  c_1 $  | 0     | $\frac{1}{2}$ | $\frac{1}{2}$ | $f_1 + \frac{3}{4}$ |
|                        | $g_1 \leq -\frac{3}{4}, f_1 \geq d - \frac{3}{4}$ | $\pm 1 + c_1$ | $2 + 2 c_1 $ | 0     | $\frac{3}{2}$ | $\frac{1}{2}$ | $f_2 + \frac{3}{4}$ |

Case (iii)  $c_3 = c_4 = 0, c_5 = \frac{1}{2}, a_3 = a_4 = \frac{1}{2}$  and  $a_5 = 0$ .

$$Q = (x_1 + a_2 x_2 + \frac{1}{2} x_3 + \frac{1}{2} x_4) x_2 - \frac{1}{2}(x_3^2 + x_4^2 + x_5^2).$$

|                        | Range  | $x_1$         | $x_1 x_2$    | $x_3$   | $x_3 x_2$ | $x_4$ | $x_5$         | $Q$                  |
|------------------------|--|---------------|--------------|---------|-----------|-------|---------------|----------------------|
|                        | $\frac{1}{8} < f_1 < d + \frac{1}{8}$              | $c_1$         | $ c_1 $      | 0       | 0         | 0     | $\frac{1}{2}$ | $f_1 - \frac{1}{8}$  |
|                        | $f_1 \geq d + \frac{1}{8}$                         | $c_1$         | $-2 c_1 $    | $\pm 1$ | 2         | 0     | $\frac{3}{2}$ | $g_2 - \frac{5}{8}$  |
| $f_1 \leq \frac{1}{8}$ | $-\frac{7}{8} < g_1 < d - \frac{7}{8}$             | $\pm 1 + c_1$ | $1 -  c_1 $  | 0       | 0         | 0     | $\frac{1}{2}$ | $g_1 + \frac{7}{8}$  |
|                        | $g_1 \leq -\frac{7}{8}, g_2 < d - \frac{27}{8}$    | $\pm 2 + c_1$ | $4 - 2 c_1 $ | $\pm 1$ | 2         | 0     | $\frac{3}{2}$ | $g_2 + \frac{27}{8}$ |
|                        | $g_1 \leq -\frac{7}{8}, g_2 \geq d - \frac{27}{8}$ | $\pm 2 + c_1$ | $4 - 2 c_1 $ | 0       | 0         | 0     | $\frac{3}{2}$ | $g_2 + \frac{23}{8}$ |
|                        | $g_1 \geq d - \frac{7}{8}$                         | $c_1$         | $2 c_1 $     | 0       | 0         | 0     | $\frac{1}{2}$ | $f_2 - \frac{1}{8}$  |

Case (iv)  $c_3 = c_4 = c_5 = 0$  and  $a_3 = a_4 = a_5 = \frac{1}{2}$ .

$$Q = (x_1 + a_2x_2 + \frac{1}{2}x_3 + \frac{1}{2}x_4 + \frac{1}{2}x_5)x_2 - \frac{1}{2}(x_3^2 + x_4^2 + x_5^2).$$

|              | Range                                   | $x_1$         | $x_2$   | $x_1x_2$      | $x_3x_2$ | $x_4x_2$ | $x_5x_2$ | $Q$                 |
|--------------|---|---------------|---------|---------------|----------|----------|----------|---------------------|
|              | $0 < f_1 < d$                           | $c_1$         | $\pm 1$ | $ c_1 $       | 0        | 0        | 0        | $f_1$               |
| $f_1 \geq d$ | $f_2 < d + 2$                           | $\pm 1 + c_1$ | $\pm 2$ | $-2 + 2 c_1 $ | 0        | 0        | 0        | $f_2 - 2$           |
|              | $f_2 \geq d + 2$                        | $c_1$         | $\pm 2$ | $-2 c_1 $     | -2       | 2        | 2        | $g_2 - \frac{1}{2}$ |
|              | $g_1 < d - 1$                           | $\pm 1 + c_1$ | $\pm 1$ | $1 -  c_1 $   | 0        | 0        | 0        | $g_1 + 1$           |
| $f_1 \leq 0$ | $g_1 > d - 1, f_2 > \frac{-1}{2}$       | $c_1$         | $\pm 2$ | $2 c_1 $      | 2        | 0        | 0        | $f_2 + \frac{1}{2}$ |
|              | $g_1 \geq d - 1, f_2 \leq \frac{-1}{2}$ | $\pm 1 + c_1$ | $\pm 2$ | $2 + 2 c_1 $  | -2       | 2        | 0        | $f_2 + 1$           |

**Lemma 12.** For  $a = A = \frac{1}{2}$ ,  $B = 0$ ,  $C = d = 1$ ,  $h_4 = h_5 = 0$ ,  $(a_3, c_3) = (0, \frac{1}{2})$  or  $(\frac{1}{2}, 0)$ ,  $(a_4, c_4) = (0, \frac{1}{2})$  or  $(\frac{1}{2}, 0)$  and  $\pm a_5 - 2c_5 \equiv 0 \pmod{1}$ , (3.10) is soluble.

*Proof.* (3.10) can be written as

$$0 < \frac{-1}{2}(x_3 - a_3x_2)^2 + [x_1 + (a_2 + \frac{1}{2}a_3^2)x_2 + a_4x_4 + a_5x_5]x_2 - \frac{1}{2}x_4^2 - x_5^2 < 1.$$

By Lemma 3, it is soluble if we can solve

$$\frac{1}{8} < [x_1 + (a_2 + \frac{1}{2}a_3^2)x_2 + a_4x_4 + a_5x_5]x_2 - \frac{1}{2}x_4^2 - x_5^2 < 1 + \frac{1}{8}.$$

It can be rewritten as

$$0 < \frac{-1}{2}(x_4 - a_4x_2)^2 + [x_1 + (a_2 + \frac{1}{2}a_3^2 + \frac{1}{2}a_4^2)x_2 + a_5x_5]x_2 - x_5^2 - \frac{1}{8} < 1,$$

which is soluble by Lemma 3, if we can solve

$$0 < [x_1 + (a_2 + \frac{1}{2}a_3^2 + \frac{1}{2}a_4^2)x_2 + a_5x_5]x_2 - x_5^2 - \frac{1}{4} < 1. \quad (5.6)$$

$$\text{Let } F = [x_1 + (a_2 + \frac{1}{2}a_3^2 + \frac{1}{2}a_4^2)x_2 + a_5x_5]x_2 - x_5^2 - \frac{1}{4}. \quad (5.7)$$

Choose  $x_5 \equiv c_5 \pmod{1}$  arbitrary,  $x_2 = \pm 1$ , then  $x_1 \equiv c_1 \pmod{1}$  such that

$$0 < F = \pm x_1 + a_2 + \frac{1}{2}a_3^2 + \frac{1}{2}a_4^2 \pm a_5x_5 - x_5^2 - \frac{1}{4} \leq |x_2| = 1.$$

If  $0 < F < 1$ , then (5.6) is soluble, otherwise we must have

$$\pm c_1 + a'_2 + a_5c_5 - c_5^2 - \frac{1}{4} \equiv 0 \pmod{1}, \quad (5.8)$$

where

$$a'_2 = a_2 + \frac{1}{2}a_3^2 + \frac{1}{2}a_4^2. \quad (5.9)$$

These congruences imply that

$$2a'_2 \equiv 2c_5^2 + \frac{1}{2} \pmod{1}. \quad (5.10)$$

Replacing  $x_1$  by  $x_1 + mx_2$  for a suitable integer  $m$ , we can suppose that  $\frac{-1}{2} < a'_2 \leq \frac{1}{2}$ . Since  $\pm a_5 - 2c_5 \equiv 0 \pmod{1}$  we have  $c_5 = 0$ ,  $\pm \frac{1}{4}$  and  $\frac{1}{2}$ . In case  $c_5 = 0$ ,  $\pm \frac{1}{4}$ , the

various cases that arise have been dealt with in the following table:

| $c_5$          | $a_5$         | $a'_2$          | $c_1$          | $(x_1, x_2, x_5)$                  | $F$             |
|----------------|---------------|-----------------|----------------|------------------------------------|-----------------|
| $\frac{1}{4}$  | $\frac{1}{2}$ | $\frac{5}{16}$  | $-\frac{1}{8}$ | $(-\frac{1}{8}, 2, \frac{1}{4})$   | $\frac{15}{16}$ |
| $\frac{1}{4}$  | $\frac{1}{2}$ | $-\frac{3}{16}$ | $\frac{3}{8}$  | $(-\frac{5}{8}, -2, -\frac{3}{4})$ | $\frac{7}{16}$  |
| $-\frac{1}{4}$ | $\frac{1}{2}$ | $\frac{5}{16}$  | $\frac{1}{8}$  | $(\frac{1}{8}, 2, -\frac{1}{4})$   | $\frac{15}{16}$ |
| $-\frac{1}{4}$ | $\frac{1}{2}$ | $-\frac{3}{16}$ | $-\frac{3}{8}$ | $(\frac{5}{8}, 2, \frac{3}{4})$    | $\frac{7}{16}$  |
| 0              | 0             | $\frac{1}{4}$   | 0              | $(0, 2, 0)$                        | $\frac{3}{4}$   |
| 0              | 0             | $-\frac{1}{4}$  | $\frac{1}{2}$  | $(\frac{3}{2}, 2, 1)$              | $\frac{3}{4}$   |

Since  $0 < F < 1$  in each case, it follows that (5.6) and hence (3.10) is soluble if  $c_5 = 0$  or  $\pm \frac{1}{4}$ . Now we can suppose that  $c_5 = \frac{1}{2}$  so that  $a_5 = 0$  and by (5.10) we have  $2a'_2 \equiv 0 \pmod{1}$ . Since  $a'_2 = a_2 + \frac{1}{2}a_3^2 + \frac{1}{2}a_4^2$ , we have  $-\frac{1}{2} < a'_2 \leq \frac{3}{4}$  so that  $a'_2 = 0$  or  $\frac{1}{2}$  and hence from (5.8)  $c_1 = \frac{1}{2}$  or 0 respectively. For  $a'_2 = 0$  we choose  $(x_1, x_2, x_5) = (\frac{1}{2}, 2, \frac{1}{2})$  to get  $F = \frac{1}{2}$ , so that (5.6) and hence (3.10) is soluble. For  $(a'_2, c_1) = (\frac{1}{2}, 0)$  by (5.9) and the hypothesis of the Lemma we must have  $(a_2, a_3, a_4) = (\frac{1}{4}, \frac{1}{2}, \frac{1}{2})$  or  $(\frac{3}{8}, \frac{1}{2}, 0)$  or  $(\frac{3}{8}, 0, \frac{1}{2})$  or  $(\frac{1}{2}, 0, 0)$ . The various cases have been dealt in the following table:

| $a_2$         | $a_3$         | $a_4$         | $a_5$ | $c_1$ | $c_2$ | $c_3$         | $c_4$         | $c_5$         | $(x_1, \dots, x_5)$                             | $Q$           |
|---------------|---------------|---------------|-------|-------|-------|---------------|---------------|---------------|---|---------------|
| $\frac{1}{4}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | 0     | 0     | 0     | 0             | 0             | $\frac{1}{2}$ | $(0, 2, 0, 0, \frac{1}{2})$                     | $\frac{3}{4}$ |
| $\frac{3}{8}$ | $\frac{1}{2}$ | 0             | 0     | 0     | 0     | 0             | $\frac{1}{2}$ | $\frac{1}{2}$ | $(1, 2, -1, \frac{3}{2}, \frac{1}{2})$          | $\frac{5}{8}$ |
| $\frac{1}{2}$ | 0             | 0             | 0     | 0     | 0     | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $(0, 2, \frac{3}{2}, \frac{1}{2}, \frac{1}{2})$ | $\frac{1}{2}$ |
| $\frac{3}{8}$ | 0             | $\frac{1}{2}$ | 0     | 0     | 0     | $\frac{1}{2}$ | 0             | $\frac{1}{2}$ | $(1, 2, \frac{3}{2}, -1, \frac{1}{2})$          | $\frac{5}{8}$ |

Since  $0 < Q < 1$ , (3.10) is soluble in each case.

**Remark 3.** Lemmas 10, 11 and 12 complete the case  $a = \frac{1}{2}$ ,  $h_4 = h_5 = 0$  ( $a_3, c_3$ ) =  $(0, \frac{1}{2})$  or  $(\frac{1}{2}, 0)$ .

## 6. $c_2 \equiv 0 \pmod{1}$ , $d \leq 1$ , $a < \frac{1}{2}$ and $a + d \leq 1$

If  $m = 1$ , then by (3.9) we have  $\frac{d}{A} \leq \frac{4d}{3a} \leq \frac{8}{3}$ , so that  $K \leq 2$ . Therefore in view of

Remark 1, inequality (4.7) and Lemma 8, we need to discuss the following cases:

- (i)  $m = 2$ ,  $K = 1$ .
- (ii)  $m = K = 1$ .
- (iii)  $m = 1$ ,  $K = 2$ .

6.1  $m = 2, K = 1$

*Lemma 13.* For  $m = 2$  and  $K = 1$ , (4.5) and hence (3.10) is soluble.

*Proof.* By definition of  $m, K$  and by (3.7) we have

$$\frac{5d}{4} = d + \frac{d}{4} \leq d + \frac{3a}{4} \leq 2A \leq 2\sqrt{d^5/3a} \leq 2d^2 \quad \text{so that } d \geq \frac{5}{8}. \quad (6.1)$$

Choose  $x_5 \equiv c_5 \pmod{1}$  arbitrary,  $x_2 = \pm 1$ . Write  $x_1 = x + c_1$  and  $x_4 = y + c_4$ . Then (4.5) reduces to an inequality of the type (2.2) with  $\alpha = A$  and  $\beta = \pm a'_4 - 2Ac_4 - Bx_5$ . Since

$$|\frac{1}{2} - A| + \frac{1}{2} < d + \frac{3a}{4},$$

taking  $h = \frac{1}{2}, k = 1$  and  $\gamma = d + \frac{3a}{4}$ , it follows from Lemma 5 that (4.5) is soluble unless  $A = \frac{1}{2}$  and  $\pm a'_4 - 2c_4 - Bx_5 \equiv \frac{1}{2} \pmod{1}$ . Taking  $x_5 = c_5$  and  $1 + c_5$  we get  $B \equiv 0 \pmod{1}$ . Since  $0 \leq B \leq A = \frac{1}{2}$  we have  $B = 0$ . If  $A = \frac{1}{2}$  and  $B = 0$  then by (3.7) and (4.3) we have

$$\frac{1}{4} = A^2 \leq AC = \det \psi = \frac{d^5}{4a} \leq \frac{3d^4}{4},$$

so that  $d > \frac{3}{4}$  which is not possible since  $a + d \leq 1$  and  $m = 2$  give  $d \leq \frac{3}{4}$ . This completes the case  $m = 2$  and  $K = 1$ .

6.2  $m = K = 1$

Here  $a + d \leq 1$  and  $a \geq \frac{d}{2}$  give  $d \leq \frac{2}{3}$ . Moreover by (4.3) we have  $d \geq \frac{1}{2}$ . Therefore

$$\frac{1}{2} \leq d \leq \frac{2}{3}. \quad (6.2)$$

*Lemma 14.* For  $m = K = 1$ , (4.5) and hence (3.10) is soluble unless  $A = B = C = \frac{1}{3}$ .

*Proof.* Using (3.7), (4.3) and (6.2) we have for  $m = K = 1$

$$\frac{d}{2} \leq A \leq \sqrt{d^5/3a} \leq \sqrt{2/3} d^2 \leq \frac{4}{9} \sqrt{2/3} < \frac{1}{2} \quad (6.3)$$

and hence

$$d \geq \sqrt{3/8}. \quad (6.4)$$

Choose  $x_5 \equiv c_5 \pmod{1}$  arbitrary,  $x_2 = \pm 1$ . Write  $x_1 = x + c_1$  and  $x_4 = y + c_4$ . Then (4.5) reduces to an inequality of the type (2.2) with  $\alpha = A$ ,  $\gamma = d$  and  $\beta = \pm a'_4 -$

$2Ac_4 - Bx_5$ . Apply Lemma 5, choosing  $h$  and  $k$  as given in the table below

| $h$           | $k$ | Range of $\alpha$                                    |
|---------------|-----|--|
| $\frac{1}{2}$ | 1   | $1 - d < A$  |
| 3             | 3   | $\frac{1}{3} < A \leq \min\{1 - d, \sqrt{2/3}d^2\}$  |
| 3             | 3   | $\frac{1}{9}(\frac{7}{2} - d) < A < \frac{1}{3}$     |
| 8             | 5   | $\frac{8}{25} < A \leq \frac{1}{9}(\frac{7}{2} - d)$ |
| 3             | 3   | $A = \frac{8}{25}$                                   |
| 8             | 5   | $\frac{1}{25}(\frac{17}{2} - d) < A < \frac{8}{25}$  |

In each case  $|h - k^2 A| + \frac{1}{2} < d$ , so that (4.5) is soluble unless  $A = \frac{1}{3}$  or  $A \leq \frac{1}{25}(\frac{17}{2} - d)$ .

If  $A \leq \frac{1}{25}(\frac{17}{2} - d)$ , then using  $A \geq \frac{d}{2}$ , we have

$$d \leq \frac{17}{27}. \quad (6.5)$$

Moreover, for  $m = K = 1$ , by (3.8)

$$t = \frac{d^5}{4aA} \leq d^3 \leq (\frac{17}{27})^3 < \frac{1}{4}. \quad (6.6)$$

We have seen that (4.5) is soluble if we can solve (4.8) which in this case reduces to

$$0 < (x_1 + a_2''x_2 + a_5''x_5)x_2 - tx_5^2 - \frac{1}{4}(a + A) < d. \quad (6.7)$$

Choose  $x_2 = \pm 1$ . Write  $x_1 = x + c_1$  and  $x_5 = y + c_5$ . Then (6.7) reduces to an inequality of the type (2.2) with  $\alpha = t$ ,  $\gamma = d$  and  $\beta = \pm a_5'' - 2tc_5$ .

Apply Lemma 5, choosing  $h$  and  $k$  as given below

| $h$ | $k$ | Range of $\alpha$                     |
|-----|-----|---------------------------------------|
| 2   | 2   | $t > \frac{1}{4}(\frac{3}{2} - d)$    |
| 2   | 3   | $t \leq \frac{1}{4}(\frac{3}{2} - d)$ |

In each case  $|h - k^2 t| + \frac{1}{2} < d$  is satisfied and  $h/k^2 \neq \alpha$ . Therefore (4.5) is soluble.

For  $A = \frac{1}{3}$ , it follows from Lemma 5 with  $h = k = 3$  that (4.5) is soluble unless

$$\beta = \pm a_4' - 2Ac_4 - Bx_5 \equiv \frac{1}{3} \pmod{\frac{1}{3}, \frac{2}{3}} \text{ or } \pm a_4' - \frac{2}{3}c_4 - Bx_5 \equiv 0 \pmod{\frac{1}{3}}.$$

Taking  $x_5 = c_5$  and  $1 + c_5$ , we get  $B \equiv 0 \pmod{\frac{1}{3}}$ . Since  $0 \leq B \leq A = \frac{1}{3}$  we have  $B = 0$

or  $\frac{1}{3}$ . If  $A = \frac{1}{3}$  and  $B = 0$ , then by (3.7)

$$\frac{1}{9} = A^2 \leq AC = \frac{C}{3} = \det \psi = \frac{d^5}{4a} \leq \frac{d^4}{2},$$

so that  $d > 0.68...$  which is not possible since  $d \leq \frac{2}{3}$  by (6.2).

Now suppose that  $A = B = \frac{1}{3}$ . By (3.7) we have

$$\frac{1}{9} = A^2 \leq AC = \frac{C}{3} \leq \frac{4}{3} \det \psi = \frac{d^5}{3a} \leq \frac{2d^4}{3}$$

or

$$\frac{1}{3} = A \leq AC \leq 2d^4 < \frac{1}{2} \quad (\text{by (6.2)}) \quad (6.8)$$

and

$$d^4 \geq \frac{1}{6}. \quad (6.9)$$

Now choose  $x_4 = c_4$ ,  $x_2 = \pm 1$ . Write  $x_1 = x + c_1$ ,  $x_5 = y + c_5$ . Then (4.5) reduces to an inequality of the type (2.2) with  $\alpha = c$ ,  $\gamma = d$  and  $\beta = \pm a'_5 - 2Cc_5 - \frac{1}{3}c_4$ . Apply Lemma 5 choosing  $h$  and  $k$  as given below:

| $h$           | $k$ | Range of $C$                                      |
|---------------|-----|---|
| $\frac{1}{2}$ | 1   | $C > 1 - d$                                       |
| $\frac{3}{2}$ | 2   | $\frac{1}{4}(2 - d) < C \leq 1 - d < \frac{3}{8}$ |
| 3             | 3   | $C \leq \frac{1}{4}(2 - d)$                       |

In each case  $|h - k^2 C| + \frac{1}{2} < d$  is satisfied, therefore (4.5) is soluble unless  $A = B = C = \frac{1}{3}$ . This proves the lemma.

**Lemma 15.** For  $A = B = C = \frac{1}{3}$ ,  $m = K = 1$ , (3.10) is soluble.

*Proof.* For  $A = B = C = \frac{1}{3}$ , we have

$$\frac{1}{12} = AC - \frac{1}{4}B^2 = \det \psi = \frac{d^5}{4a},$$

or

$$\frac{d}{2} \leq a = 3d^5 \quad \text{so that } d^4 \geq \frac{1}{6}. \quad (6.10)$$

Choose  $(x_4, x_5) \equiv (c_4, c_5) \pmod{1}$  arbitrary,  $x_2 = \pm 1$ . Write  $x_1 = x + c_1$  and  $x_3 = y + c_3$ . Then (3.10) reduces to an inequality of the type (2.2) with  $\alpha = a$ ,  $\gamma = d$  and  $\beta = \pm a_3 - 2a(c_3 + h_4x_4 + h_5x_5)$ . Taking  $h = k = 3$  it follows from Lemma 5, that (3.10) is soluble unless  $a = \frac{1}{3}$  and  $\beta \equiv h/k \pmod{1/k, 2\alpha}$  or

$$\pm a_3 - 2(c_3 + h_4x_4 + h_5x_5)/3 \equiv 0 \pmod{\frac{1}{3}}.$$

Taking  $x_4 = c_4$  and  $1 + c_4$ , we get  $2h_4 \equiv 0 \pmod{1}$ . Since  $-\frac{1}{2} < h_4 \leq \frac{1}{2}$ , we have  $h_4 = 0$  or  $\frac{1}{2}$ . Similarly  $h_5 = 0$  or  $\frac{1}{2}$ .



In case  $A = B = C = \frac{1}{3}$  and  $h_4 = \frac{1}{2}$ , then (3.10) can be written as

$$0 < (x_1 + \dots)x_2 - \frac{5}{12}x_4^2 - \frac{1}{3}[(x_3 + h_5x_5)^2 + x_3x_4 + (1 + h_5)x_5x_4 + x_5^2] < d. \quad (6.11)$$

Choose  $(x_3, x_5) = (c_3, c_5)$ ,  $x_2 = \pm 1$ ,  $x_1 = x + c_1$  and  $x_4 = y + c_4$ . Then (6.11) reduces to an inequality of the type (2.2) with  $\alpha = \frac{5}{12}$  and  $\gamma = d$ . Since  $|\frac{1}{2} - \frac{5}{12}| + \frac{1}{2} = \frac{7}{12} < d$ , (6.11) is soluble by Lemma 5 with  $h = \frac{1}{2}$  and  $k = 1$ .

Now we can suppose that  $h_4 = 0$ . Similarly  $h_5 = 0$ , so that (3.10) reduces to

$$0 < (x_1 + \dots)x_2 - \frac{1}{3}(x_3^2 + x_4^2 + x_4x_5 + x_5^2) < d,$$

which can be written as

$$0 < \frac{-1}{3}(x_4 + \frac{1}{2}x_5 + \lambda x_2)^2 + (x_1 + a'_2x_2 + \dots)x_2 - \frac{1}{3}x_3^2 - \frac{1}{4}x_5^2 < d. \quad (6.12)$$

Since  $1 < 3d < 2$ , (6.12) is soluble by Lemma 3, if we can solve

$$\frac{1}{12} < (x_1 + \dots)x_2 - \frac{1}{3}x_3^2 - \frac{1}{4}x_5^2 < d + \frac{1}{12},$$

or if we can solve

$$0 < \frac{-1}{4}(x_5 + \lambda_1x_2)^2 + (x_1 + a''_2x_2 + \dots)x_2 - \frac{1}{3}x_3^2 - \frac{1}{12} < d.$$

Since  $2 < 4d < 3$ , it is soluble by Lemma 3, if we can solve

$$\frac{1}{16} < (x_1 + \dots)x_2 - \frac{1}{3}x_3^2 - \frac{1}{12} < d + \frac{1}{4},$$

or

$$0 < (x_1 + \dots)x_2 - \frac{1}{3}x_3^2 - \frac{7}{48} < d + \frac{3}{16}.$$

Choose  $x_2 = 1$ ,  $x_1 = x + c_1$  and  $x_3 = y + c_3$ , then the above inequality reduces to an inequality of the type (2.2) with  $\alpha = \frac{1}{3}$  and  $\gamma = d + \frac{3}{16}$ . Since  $|\frac{1}{2} - \frac{1}{3}| + \frac{1}{2} = \frac{2}{3} < d + \frac{3}{16}$ , it follows from Lemma 5, with  $h = \frac{1}{2}$  and  $k = 1$ , that the above inequality is soluble. This proves the lemma.

Lemmas 14 and 15 complete the case  $m = K = 1$ .

### 6.3 $m = 1, K = 2$

In this case (4.8) reduces to

$$0 < (x_1 + a''_2x_2 + a''_5x_5)x_2 - tx_5^2 - \frac{1}{4}(a + A) < d + \frac{3}{4}A. \quad (6.13)$$

By (3.8) and definition of  $m$  and  $K$

$$t = \frac{d^5}{4aA} \leq \frac{4d^3}{3}, \quad (6.14)$$

so that by (6.2) and (3.9) we have

$$\frac{1}{t} \left( d + \frac{3}{4}A \right) \geq \frac{3}{4d^3} \left[ d + \frac{9d}{32} \right] = \frac{123}{128d^2} \geq \frac{123}{128} \left( \frac{9}{4} \right) > 2.$$

Let  $L$  be the integer satisfying

$$L < \frac{1}{t} \left( d + \frac{3A}{4} \right) \leq L + 1, \quad (6.15)$$

then  $L \geq 2$ . Now (6.13) can be rewritten as

$$0 < -t(x_5 + \lambda x_2)^2 + (x_1 + a_2^* x_2)x_2 - \frac{1}{4}(a + A) < d + \frac{3}{4}A,$$

which is soluble by Lemma 3, if we can solve

$$\frac{t}{4} < (x_1 + \dots)x_2 - \frac{1}{4}(a + A + t) < d + \frac{3A}{4} + \frac{1}{4}(L^2 - 1)t. \quad (6.16)$$

Choosing  $x_2 = 1$  and  $x_1 \equiv c_1 \pmod{1}$  in such a way that

$$0 < (x_1 + \dots)x_2 - \frac{1}{4}(a + A + t) \leq |x_2| = 1,$$

we see that (6.16) is soluble if  $d + \frac{3A}{4} + \frac{1}{4}(L^2 - 1)t > 1$ . In particular, it is soluble for  $L \geq 4$ . Therefore we can suppose that  $L = 2$  or  $3$  and

$$d + \frac{3A}{4} + \frac{1}{4}(L^2 - 1)t \leq 1, \quad (6.17)$$

and hence

$$\begin{aligned} 1 &\geq d + \frac{3A}{4} + \frac{1}{4}(L^2 - 1)t \geq \left( d + \frac{3A}{4} \right) \left( \frac{L+3}{4} \right) \geq \left( d + \frac{9d}{32} \right) \left( \frac{L+3}{4} \right) \\ &= \frac{41d(L+3)}{128}, \end{aligned} \quad (6.18)$$

or

$$d \leq \frac{128}{41(L+3)}.$$

Moreover by (3.9), (4.3), (6.14) and (6.15), we have

$$\frac{41d}{32} \leq d + \frac{3A}{4} \leq (L+1)t \leq \frac{4(L+1)d^3}{3}, \quad (6.19)$$

so that

$$d^2 \geq \frac{123}{128(L+1)}. \quad (6.20)$$

Choose  $x_2 = \pm 1$ . Write  $x_1 = x + c_1$  and  $x_5 = y + c_5$ . Then (6.13) reduces to an inequality of the type (2.2) with  $\alpha = t$ ,  $\gamma = d + \frac{3A}{4}$  and  $\beta = \pm a_5'' - 2tc_5$ . Apply Lemma 5, with  $(h, k) = (1, 2)$  or  $(\frac{3}{2}, 3)$  according as  $L = 2$  or  $3$  respectively. In both the cases

$$|h - k^2 t| + \frac{1}{2} < d + \frac{3A}{4},$$

is satisfied, therefore (6.13) is soluble unless

$$\left. \begin{array}{l} \text{(i) } L=2, t=\frac{1}{4} \text{ and } \beta = \pm a_5'' - 2tc_5 \equiv 0 \pmod{\frac{1}{2}} \text{ or} \\ \text{(ii) } L=3, t=\frac{1}{6} \text{ and } \pm a_5'' - 2tc_5 \equiv \frac{1}{2} \pmod{\frac{1}{3}}. \end{array} \right\} \quad (6.21)$$

*Lemma 16.* For  $L=2, t=\frac{1}{4}$  and  $\pm a_5'' - 2tc_5 \equiv 0 \pmod{\frac{1}{2}}$ , (4.5) is soluble.

*Proof.* For  $t=\frac{1}{4}$ , by (6.14) and (6.15) we have

$$d + \frac{3A}{4} \leq \frac{3}{4} \quad \text{and} \quad d^3 \geq \frac{3}{16} \quad \text{or} \quad d > \frac{4}{7}. \quad (6.22)$$

Moreover  $A = \frac{d^5}{a} \leq 2d^4$ . Choose  $x_5 = c_5, x_2 = 1, x_1 = x + c_1$  and  $x_4 = y + c_4$ . Then (4.5) reduces to an inequality of the type (2.2) with  $\alpha = A$  and  $\gamma = d$ . Apply Lemma 5 with  $(h, k) = (2, 3)$  if  $A \neq \frac{2}{9}$  and  $(h, k) = (\frac{7}{2}, 4)$  if  $A = \frac{2}{9}$ . In each case  $\alpha \neq h/k^2$  and  $|h - k^2 A| + \frac{1}{2} < d$ , so that (4.5) is soluble. This proves the Lemma.

*Lemma 17.* For  $L=3, t=\frac{1}{6}$  and  $\pm a_5'' - \frac{1}{3}c_5 \equiv \frac{1}{2} \pmod{\frac{1}{3}}$ , (6.13) is soluble.

*Proof.* For  $L=3$  and  $t=\frac{1}{6}$ , (6.13) becomes

$$0 < (x_1 + a_2''x_2 + a_5''x_5)x_2 - \frac{1}{6}x_5^2 - \frac{1}{4}(a + A) < d + \frac{3A}{4}. \quad (6.23)$$

Let  $F = (x_1 + a_2''x_2 + a_5''x_5)x_2 - \frac{1}{6}x_5^2$ . Replacing  $F$  by an equivalent form we can suppose that  $-\frac{1}{2} < a_5'' \leq \frac{1}{2}$  so that  $\pm a_5'' - \frac{1}{3}c_5 \equiv \frac{1}{2} \pmod{\frac{1}{3}}$  implies that either

- (i)  $c_5 = 0$  and  $a_5'' = \pm \frac{1}{6}, \frac{1}{2}$  or
- (ii)  $c_5 = \frac{1}{2}$  and  $a_5'' = 0, \pm \frac{1}{3}$ .

Replacing  $x_5$  by  $x_2 \pm x_5$  if  $a_5'' = \pm \frac{1}{3}$  and by  $x_2 \mp x_5$  if  $a_5'' = \pm \frac{1}{6}$  it can be easily seen that (6.23) is soluble unless  $(a_5'', c_5) = (0, \frac{1}{2})$  or  $(\frac{1}{2}, 0)$ .

*Remark 4.* Now we are left with  $t=\frac{1}{6}, (a_5'', c_5) = (0, \frac{1}{2})$  or  $(\frac{1}{2}, 0)$  in which case by (3.8) we have

$$\frac{1}{6} = t = \frac{d^5}{4aA} \quad \text{or} \quad A = \frac{3d^5}{2a} \leq 3d^4, \quad (6.24)$$

so that from (6.15) and (3.5)

$$\frac{2}{3} \geq d + \frac{3}{4}A = d + \frac{9d^5}{8a} \geq d + \frac{9}{8}d^5(2/d^5)^{1/3} = d + \frac{9}{8}(2d^5)^{1/3}$$

and hence

$$d \leq 0.514 \dots \text{ and } A < \frac{3}{14}. \quad (6.25)$$

Moreover,  $d \geq \frac{1}{2}$  by (6.2).

$$0 < f = (x_1 + a_2''x_2)x_2 - \frac{1}{6}x_5^2 - \frac{1}{4}(a + A) < d + \frac{3A}{4}. \quad (6.26)$$

Replacing  $x_1$  by  $x_1 + nx_2$  for a suitable integer  $n$  we can suppose that  $-\frac{1}{2} < a_2'' \leq \frac{1}{2}$ . Depending on the values of  $c_1$  and  $a_2''$ , the inequality (6.26) can be satisfied for suitable choice of  $x_1, x_2$  and  $x_5$  as given below: Let  $f_n = n|c_1| + n^2a_2''$ ,  $g_n = -n|c_1| + n^2a_2''$

| Range  | $x_1$         | $x_1x_2$     | $x_5$         | $f$  |
|--|---------------|--------------|---------------|--|
| $\frac{1}{24} + \frac{1}{4}(a + A) < f_1 < \frac{1}{24} + \frac{a}{4} + d + A$                 | $c_1$         | $ c_1 $      | $\frac{1}{2}$ | $f_1 - \frac{1}{24} - \frac{1}{4}(a + A)$  |
| $f_1 \geq \frac{1}{24} + \frac{a}{4} + d + A$  | $c_1$         | $ c_1 $      | $\frac{3}{2}$ | $f_1 - \frac{3}{8} - \frac{1}{4}(a + A)$   |
| $f \leq \frac{1}{24} + \frac{1}{4}(a + A)$ ,<br>$f_1 < -\frac{5}{8} + d + A + \frac{a}{4}$     | $\pm 1 + c_1$ | $1 +  c_1 $  | $\frac{3}{2}$ | $f_1 - \frac{5}{8} - \frac{1}{4}(a + A)$   |
| $f_1 \leq \frac{1}{24} + \frac{1}{4}(a + A)$ , $f_1 \geq -\frac{5}{8} + d + A + \frac{a}{4}$ , |               |              |               |  |
| $-\frac{5}{8} + d + A + \frac{a}{4} > g_1 > -\frac{5}{8} + \frac{1}{4}(a + A)$                 | $\pm 1 + c_1$ | $1 -  c_1 $  | $\frac{3}{2}$ | $g_1 + \frac{5}{8} - \frac{1}{4}(a + A)$   |
| $-\frac{23}{24} + \frac{1}{4}(a + A) < g_1 \leq -\frac{5}{8} + \frac{1}{4}(a + A)$             | $\pm 1 + c_1$ | $1 -  c_1 $  | $\frac{1}{2}$ | $g_1 + \frac{23}{24} - \frac{1}{4}(a + A)$ |
| $g_1 \leq -\frac{23}{24} + \frac{1}{4}(a + A)$   | $\pm 1 + c_1$ | $2 + 2 c_1 $ | $\frac{5}{2}$ | $f_2 + \frac{23}{24} - \frac{1}{4}(a + A)$ |
| $g_1 \geq -\frac{5}{8} + d + A + \frac{a}{4}$  | $c_1$         | $2 c_1 $     | $\frac{1}{2}$ | $f_2 - \frac{1}{24} - \frac{1}{4}(a + A)$  |

This proves the Lemma.

**Lemma 19.** For  $t = \frac{1}{6}$  and  $(a_5'', c_5) = (\frac{1}{2}, 0)$ , (4.8) is soluble unless  $c_1 = \frac{1}{2}$ ,  $d = \frac{1}{2}$ ,  $a = \frac{1}{4}$  and  $A = \frac{3}{16}$ .

*Proof.* In this case (4.8) can be written as

$$0 < f = (x_1 + a_2''x_2 + \frac{1}{2}x_5)x_2 - \frac{1}{6}x_5^2 - \frac{1}{4}(a + A) < d + \frac{3}{4}A. \quad (6.27)$$

Replacing  $x_1$  by  $x_1 + nx_2$  for a suitable integer  $n$  we can suppose that  $-\frac{1}{2} < a_2'' \leq \frac{1}{2}$ . Depending on the values of  $c_1$  and  $a_2''$  for suitable choice of  $x_1, x_2$  and  $x_5$  as given below, the inequality (6.27) can be satisfied unless

$$f_1 \geq d + A + \frac{a}{4}, g_1 \leq -\frac{1}{3} + \frac{1}{4}(a + A) \text{ and } -6 + g_6 \leq \frac{1}{4}(a + A) \text{ where}$$

$$f_n = n|c_1| + n^2a_2'' \text{ and } g_1 = -n|c_1| + n^2a_2''.$$

Now we suppose that  $-6 + g_6 \leq \frac{1}{4}(a + A)$ ,  $f_1 \geq d + A + \frac{a}{4}$  and  $g_1 \leq -\frac{1}{3} + \frac{1}{4}(a + A)$  so that

$$a_2'' \geq d + A + \frac{a}{4} - |c_1| \geq \frac{1}{2} + \frac{3}{4}a + \frac{a}{4} - \frac{1}{2} = a \geq \frac{d}{2} \geq \frac{1}{4}. \quad (6.28)$$

Let  $|c_1| = \frac{1}{2} - \varepsilon_1$ ,  $a_2'' = \frac{1}{4} + \varepsilon_2$ ,  $\varepsilon_1, \varepsilon_2 \geq 0$ .

| Range  |               |         |               |         |           |   |
|--|---------------|---------|---------------|---------|-----------|---|
|  | $x_1$         | $x_2$   | $x_1 x_2$     | $x_5$   | $x_2 x_5$ | $f$                                     |
| $\frac{1}{4}(a+A) < f_1 < d+A+\frac{a}{4}$     | $c_1$         | $\pm 1$ | $ c_1 $       | 0       | 0         | $f_1 - \frac{1}{4}(a+A)$                |
| $f_1 \leq \frac{1}{4}(a+A)$                    |               |         |               |         |           |   |
| $f_1 > -\frac{1}{3} + \frac{1}{4}(a+A)$        | $c_1$         | $\pm 1$ | $ c_1 $       | $\pm 1$ | 1         | $f_1 + \frac{1}{3} - \frac{1}{4}(a+A)$  |
| $f_1 \leq -\frac{1}{3} + \frac{1}{4}(a+A),$    | $\pm 1 + c_1$ | $\pm 1$ | $1 -  c_1 $   | 0       | 0         | $g_1 + 1 - \frac{1}{4}(a+A)$            |
| $g_1 < d+A+\frac{a}{4}-1.$                     |               |         |               |         |           |   |
| $f_1 \leq -\frac{1}{3} + \frac{1}{4}(a+A),$    | $c_1$         | $\pm 2$ | $2 c_1 $      | $\pm 2$ | 4         | $f_2 + \frac{4}{3} - \frac{1}{4}(a+A)$  |
| $g_1 \geq d+A+\frac{a}{4}-1.$                  |               |         |               |         |           |   |
| $f_1 \geq -\frac{1}{3} + \frac{1}{4}(a+A)$     | $c_1$         | $\pm 1$ | $- c_1 $      | $\pm 1$ | 1         | $g_1 + \frac{1}{3} - \frac{1}{4}(a+A)$  |
| $d+A+\frac{a}{4}$                              |               |         |               |         |           |   |
| $g_1 \leq -\frac{1}{3} + \frac{1}{4}(a+A)$ and |               |         |               |         |           |   |
| $g_2 > \frac{1}{4}(a+A)$                       | $c_1$         | $\pm 2$ | $-2 c_1 $     | 0       | 0         | $g_2 - \frac{1}{4}(a+A)$                |
| $g_2 \leq \frac{1}{4}(a+A),$                   | $\pm 1 + c_1$ | $\pm 4$ | $-4 - 4 c_1 $ | $\pm 1$ | 4         | $g_4 - \frac{13}{6} - \frac{1}{4}(a+A)$ |
| $g_4 > \frac{13}{6} + \frac{1}{4}(a+A)$        |               |         |               |         |           |   |
| $g_2 \leq \frac{1}{4}(a+A),$                   | $\pm 1 + c_1$ | $\pm 6$ | $-6 - 6 c_1 $ | 0       | 0         | $g_6 - 6 - \frac{1}{4}(a+A)$            |
| $g_4 \leq \frac{13}{6} + \frac{1}{4}(a+A)$     |               |         |               |         |           |   |

Therefore  $\frac{1}{4}(a+A) \geq -6 + g_6 = 6\varepsilon_1 + 36\varepsilon_2$  and hence

$$8\varepsilon_1 + 64\varepsilon_2 < 12\varepsilon_1 + 72\varepsilon_2 < \frac{1}{2}(a+A) < \frac{1}{6} + \frac{1}{4}(a+A). \quad (6.29)$$

Suppose now that  $(\varepsilon_1, \varepsilon_2) \neq (0, 0)$ . Choose an integer  $m$  such that

$$4m\varepsilon_1 + 16m^2\varepsilon_2 \leq \frac{1}{6} + \frac{1}{4}(a+A) < 4(m+1)\varepsilon_1 + 16(m+1)^2\varepsilon_2. \quad (6.30)$$

By (6.29) we have  $m \geq 2$ . Choosing  $x_1 = \pm(m+1) + c_1$ ,  $x_2 = \pm 4(m+1)$ , so that  $x_1 x_2 = -4(m+1)^2 - 4(m+1)|c_1|$ ,  $x_5 = \pm 1$  so that  $x_2 x_5 = 4(m+1)$ , we have by (6.30)

$$\begin{aligned} 0 < f &= -4(m+1)^2 - 4(m+1)|c_1| + 16(m+1)^2 a_2'' + 2(m+1) - \frac{1}{6} - \frac{1}{4}(a+A) \\ &= 4(m+1)\varepsilon_1 + 16(m+1)^2\varepsilon_2 - \frac{1}{6} - \frac{1}{4}(a+A) \\ &< [(m+1)^2/m^2] \left[ \frac{1}{6} + \frac{1}{4}(a+A) \right] - \frac{1}{6} - \frac{1}{4}(a+A) \leq \frac{5}{4} \left( \frac{1}{6} + \frac{7A}{12} \right) \\ &< d + \frac{3A}{4}. \end{aligned}$$

Therefore (6.27) is soluble unless  $(\varepsilon_1, \varepsilon_2) = (0, 0)$  that is  $c_1 = \frac{1}{2}$  and  $a_2'' = \frac{1}{4}$ . For  $a_2'' = \frac{1}{4}$ , by (6.28) we have  $d = \frac{1}{2}$ ,  $a = \frac{d}{2} = \frac{1}{4}$  and  $A = \frac{3}{4}a = \frac{3}{16}$ . This completes the lemma.

**Lemma 20.** For  $t = \frac{1}{6}$ ,  $(a_5'', c_5) = (\frac{1}{2}, 0)$ ,  $c_1 = \frac{1}{2}$ ,  $a = \frac{1}{4}$ ,  $d = \frac{1}{2}$  and  $A = \frac{3}{16}$ , (3.10) is soluble unless  $Q$  is equivalent to  $(x_1 + \dots)x_2 - \frac{1}{4}(x_3^2 + x_4^2 + x_5^2 + x_3x_4 + x_3x_5 + x_4x_5)$  and  $(c_1, \dots, c_5)$  is equivalent to  $(\frac{1}{2}, 0, \dots, 0)$ .

*Proof.* By (3.7) we have

$$AC = \frac{3C}{16} \leq \frac{d^5}{3a} = \frac{4d^5}{3} = \frac{1}{24} \quad \left( d = \frac{1}{2} \right)$$

so that

$$\frac{9}{4} \leq \frac{d}{C} = \frac{1}{2C} \leq \frac{8}{3}.$$

(3.10) can be written as

$$0 < -C[x_1 + \dots]^2 + (x_1 + a_2^*x_2 + \dots)x_2 - \left( \frac{3}{16} - \frac{B^2}{4C} \right)x_4^2 - \frac{1}{16} < \frac{1}{2},$$

which is soluble by Lemma 3 if we can solve

$$\frac{C}{4} < (x_1 + \dots)x_2 - \left( \frac{3}{16} - \frac{B^2}{4C} \right)x_4^2 - \frac{1}{16} < \frac{1}{2} + C,$$

or

$$0 < (x_1 + \dots)x_2 - \left( \frac{3}{16} - \frac{B^2}{4C} \right)x_4^2 - \frac{1}{16} - \frac{C}{4} < \frac{1}{2} + \frac{3C}{4}. \quad (6.31)$$

Choose  $x_2 = \pm 1$ , write  $x_1 = x + c_1$ ,  $x_4 = y + c_4$ , then (6.31) reduces to an inequality of the type (2.2) with  $\alpha = \frac{3}{16} - \frac{B^2}{4C}$  and  $\gamma = \frac{1}{2} + \frac{3C}{4}$ .

Now

$$\frac{3}{16} - \frac{B^2}{4C} = \frac{1}{C} \left( \frac{3C}{16} - \frac{B^2}{4} \right) = \frac{1}{C} \det \psi = \frac{At}{C} = \frac{A}{6C} = \frac{1}{32C}, \quad (6.32)$$

so that

$$\frac{9}{64} < \frac{3}{16} - \frac{B^2}{4C} < \frac{1}{6}.$$

Apply Lemma 5 with  $h$  and  $k$  as given:

| $h$           | $k$ | Range of $\alpha$                                     |
|---------------|-----|---|
| $\frac{3}{2}$ | 3   | $\alpha > \frac{1}{9}(\frac{3}{2} - \frac{3C}{4})$    |
| $\frac{1}{2}$ | 2   | $\alpha \leq \frac{1}{9}(\frac{3}{2} - \frac{3C}{4})$ |

In each case  $|h - k^2\alpha| + \frac{1}{2} < \frac{1}{2} + \frac{3C}{4}$  is satisfied, therefore (6.31) is soluble unless  $\alpha = \frac{1}{6}$

or  $C = \frac{3}{16}$ , so that by (6.32),  $B = \frac{1}{8}$  and

$$Q = (x_1 + \dots)x_2 - \frac{1}{4}(x_3 + h_4x_4 + h_5x_5)^2 - \frac{3}{16}x_4^2 - \frac{1}{8}x_4x_5 - \frac{3}{16}x_5^2.$$

Since  $ah_4^2 + A = \frac{1}{4}h_4^2 + \frac{3}{16}$  is a value of  $\phi$  and  $a = \frac{1}{4}$  is the minimum value we have  $h_4 = \frac{1}{2}$ . Similarly  $h_5 = \frac{1}{2}$ . By symmetry  $c_3 = c_4 = c_5 = 0$  and hence the Lemma follows.

**Lemma 21.** For  $Q = (x_1 + \dots)x_2 - \frac{1}{4}(x_3^2 + x_4^2 + x_5^2 + x_3x_4 + x_3x_5 + x_4x_5)$  and  $(c_1, \dots, c_5) = (\frac{1}{2}, 0, \dots, 0)$ , (3.10) is soluble.

*Proof.* (3.10) can be written as

$$0 < (x_1 + \dots)x_2 - \frac{1}{4}(x_3^2 + x_4^2 + x_5^2 + x_3x_4 + x_3x_5 + x_4x_5) < \frac{1}{2}. \quad (6.33)$$

Choose  $x_5 = 0$ ,  $x_2 = 1$  and  $x_4 \equiv c_4 \pmod{1}$  arbitrary. Write  $x_1 = x + \frac{1}{2}$  and  $x_3 = y$ . Then (6.33) reduces to an inequality of the type (2.2) with  $\alpha = \frac{1}{4}$ ,  $\gamma = \frac{1}{2}$ ,  $\beta = a_3 - \frac{1}{4}x_4$  and  $v = \frac{1}{2} + a_2 + a_4x_4 - \frac{1}{4}x_4^2$ . Write  $x = n^2$  and  $y = 2n$ , then (6.33) is soluble if we can find an integer such that

$$0 < 2\beta n + v < \frac{1}{2}. \quad (6.34)$$

Choose

$$x_4 = \begin{cases} \pm 1 & \text{if } a_3 = 0 \\ 1 & \text{if } 1/4 < a_3 \leq 1/2, \\ 0 & \text{if } 0 < |a_3| \leq 1/4, \\ -1 & \text{if } -1/2 < a_3 < -1/4, \end{cases}$$

so that  $0 < |2\beta| < \frac{1}{2}$  except when  $a_3 = 0$ ,  $\pm \frac{1}{4}$  or  $\frac{1}{2}$ , in which case  $2\beta = \pm \frac{1}{2}$  and

$$v = \begin{cases} 1/4 + a_2 \pm a_4 & \text{if } a_3 = 0 \\ 1/2 + a_2 & \text{if } a_3 = \pm 1/4 \\ 1/4 + a_2 + a_4 & \text{if } a_3 = 1/2. \end{cases}$$

Choosing  $n$  in such a way that  $0 < 2\beta n + v < |2\beta| \leq \frac{1}{2}$ , we see that (6.34) is soluble unless  $2\beta = \pm \frac{1}{2}$  and  $v \equiv 0 \pmod{2\beta}$  or  $v \equiv 0 \pmod{\frac{1}{2}}$ . Now suppose that

$$2\beta = \pm \frac{1}{2} \text{ and } v \equiv 0 \pmod{\frac{1}{2}} \text{ or}$$

$$a_3 = 0, \pm \frac{1}{4} \text{ or } \frac{1}{2} \text{ and } v \equiv 0 \pmod{\frac{1}{2}}. \quad (6.35)$$

Case (i)  $a_3 = \pm \frac{1}{4}$ .

By (6.34) and (6.35), we have  $\frac{1}{2} + a_2 \equiv 0 \pmod{\frac{1}{2}}$  so that  $a_2 = 0$  or  $\frac{1}{2}$ . If  $a_2 = 0$ , then by Birch reduction  $a_3 = a_4 = a_5 = 0$  which is not the case. If  $a_2 = \frac{1}{2}$  then choosing  $(x_1, \dots, x_5) = (\frac{-1}{2}, 2, \mp 1, 0, 0)$ , we have  $Q = \frac{1}{4}$  so that (6.33) is soluble.

Remark 5. (6.33) is soluble in case  $a_3 = \pm \frac{1}{4}$ . Similarly we can show that it is soluble if  $a_4 = \pm \frac{1}{4}$  or  $a_5 = \pm \frac{1}{4}$ .

Case (ii)  $a_3 = 0$  or  $\frac{1}{2}$ .

By symmetry  $a_4 = 0$  or  $\frac{1}{2}$  and  $a_5 = 0$  or  $\frac{1}{2}$ .

If  $a_3 = \frac{1}{2}$ , then

$$\begin{aligned} Q &= (x_1 + a_2 x_2 + \frac{1}{2} x_3 + a_4 x_4 + a_5 x_5) x_2 - \frac{1}{4} (x_3^2 + x_4^2 + x_5^2 + x_3 x_4 + x_3 x_5 + x_4 x_5) \\ &= [x_1 + (a_2 + \frac{1}{4}) x_2 + (a_4 - \frac{1}{4}) x_4 + (a_5 - \frac{1}{4}) x_5] x_2 - \frac{1}{4} [(x_3 - x_2)^2 + x_4^2 \\ &\quad + x_5^2 + (x_3 - x_2) x_4 + (x_3 - x_2) x_5 + x_4 x_5] \\ &\sim [x_1 + a_2^* x_2 + a_4^* x_4 + a_5^* x_5] x_2 - \frac{1}{4} [x_3^2 + x_4^2 + x_5^2 + x_3 x_4 + x_3 x_5 + x_4 x_5] = Q^*. \end{aligned}$$

Since  $a_4 = 0$  or  $\frac{1}{2}$ , we have  $a_4^* = \pm \frac{1}{4}$ , therefore replacing  $Q$  by  $Q^*$  result follows from Remark 5. Thus  $a_3 = 0$  and by symmetry  $a_4 = a_5 = 0$ . By (6.34) and (6.35), we have  $\frac{1}{4} \pm a_2 + a_4 = \frac{1}{4} + a_2 \equiv 0 \pmod{\frac{1}{2}}$  so that  $a_2 = \pm \frac{1}{4}$  and choosing  $(x_1, \dots, x_5) = (\pm \frac{1}{2}, \pm 1, 0, 0, 0)$ , we have  $Q = \frac{1}{4}$  so that (6.33) is soluble. This proves the lemma.

Lemmas 1–21 complete the proof of the Theorem.

## References

- [1] Aggarwal S K and Gupta D P, Positive values of inhomogeneous quadratic forms of signature  $(-2)$ , *J. Number Theory* **29** (1988) 138–165



- [2] Aggarwal S K and Gupta D P, Least positive values of inhomogeneous quadratic forms of signature  $(-3)$ , *J. Number Theory* **37** (1991) 260–278
- [3] Aggarwal S K and Gupta D P, Positive values of inhomogeneous quadratic forms of signature  $(4)$ , *J. Indian Math. Soc.* **57** (1991) 1–23
- [4] Bambah R P, Dumir V C and Hans-Gill R J, Positive values of non-homogeneous indefinite quadratic forms, *Proc. Col. in classical number theory*, Budapest (1981) 111–170
- [5] Bambah R P, Dumir V C and Hans-Gill R J, On a conjecture of Jackson on non-homogeneous quadratic forms, *J. Number Theory* **16** (1983) 403–419
- [6] Bambah R P, Dumir V C and Hans-Gill R J, Positive values of non-homogeneous indefinite quadratic forms II, *J. Number Theory* **18** (1984) 313–341
- [7] Barnes E S, The positive values of inhomogeneous ternary quadratic forms, *J. Aust. Math. Soc.* **2** (1961) 127–132
- [8] Birch B J, The inhomogeneous minimum of quadratic forms of signature zero, *Acta* **3** (1958) 85–98
- [9] Blaney H, Indefinite quadratic forms in  $n$ -variables, *J. London Math. Soc.* **23** (1948) 153–160
- [10] Blaney H, Indefinite ternary quadratic forms, *Quart. J. Math. Oxford* **1** (1950) 262–269
- [11] Davenport H and Heilbronn H, Asymmetric inequalities for non-homogeneous linear forms, *J. London Math. Soc.* **22** (1947) 53–61
- [12] Dumir V C, Asymmetric inequalities for non-homogeneous ternary quadratic forms, *Proc. Cambridge Philos. Soc.* **63** (1967) 291–303
- [13] Dumir V C, Positive values of inhomogeneous quadratic forms I, *J. Aust. Math. Soc.* **8** (1968) 87–101
- [14] Dumir V C, Positive values of inhomogeneous quadratic forms II, *J. Aust. Math. Soc.* **8** (1968) 287–303
- [15] Dumir V C and Hans-Gill R J, On positive values of non-homogeneous quaternary quadratic forms of type  $(1, 3)$ , *Indian J. Pure Appl. Math.* **12** (1981) 814–825
- [16] Dumir V C and Sehmi Ranjeet, Positive values of non-homogeneous indefinite quadratic forms of type  $(2, 5)$ , *J. Number Theory* (to appear)
- [17] Hans-Gill R J and Raka Madhu, Positive values of inhomogeneous 5-ary quadratic forms of type  $(3, 2)$ , *J. Aust. Math. Soc.* **29** (1980) 439–450
- [18] Hans-Gill R J and Raka Madhu, Positive values of inhomogeneous quinary quadratic forms of type  $(4, 1)$ , *J. Aust. Math. Soc.* **31** (1981) 175–188
- [19] Jackson T H, Gaps between values of quadratic forms, *J. London Math. Soc.* **3** (1971) 47–58
- [20] Macbeath A M, A new sequence of minima in the geometry of numbers, *Proc. Cambridge Philos. Soc.* **47** (1951) 266–273
- [21] Marguils G A, Indefinite quadratic forms and unipotent flows on homogeneous spaces, *Comp. Rend. Acad. Sci.* **304** (1987) 249–253
- [22] Watson G L, Indefinite quadratic polynomials, *Mathematika* **7** (1960) 141–144



## Extended Kac–Akhiezer formulae and the Fredholm determinant of finite section Hilbert–Schmidt kernels

S GANAPATHI RAMAN and R VITTAL RAO

Department of Mathematics, Indian Institute of Science, Bangalore 560 012, India

MS received 29 October 1993; revised 15 February 1994

**Abstract.** This paper deals with some results (known as Kac–Akhiezer formulae) on generalized Fredholm determinants for Hilbert–Schmidt operators on  $L_2$ -spaces, available in the literature for convolution kernels on intervals. The Kac–Akhiezer formulae have been obtained for kernels which are not necessarily of convolution nature and for domains in  $\mathbb{R}^n$ .

**Keywords.** Hilbert–Schmidt integral operator; Fredholm determinant; Kac–Akhiezer formula.

### 1. Introduction

The classical Fredholm determinant [2] of a given symmetric Hilbert–Schmidt integral (briefly H-S) operator  $T$  is an analytic function  $D_T(\lambda)$  with the property that  $\lambda (\neq 0)$  is a zero of  $D_T(\lambda)$  if and only if  $1/\lambda$  is an eigenvalue of  $T$ . We shall call any analytic function  $f_T(\lambda)$ , as a Fredholm determinant of  $T$  if its zeros are precisely reciprocals of the non-zero eigenvalues of  $T$ .

For convolution kernels such a Fredholm determinant was obtained by Kac [3], [4] and Akhiezer [1]. Their result is briefly summarized below:

Let  $k(x)$  satisfy the following conditions:

1.  $k(x)$  is a bounded, continuous, real valued function on  $(-\infty, \infty)$ ,
2.  $k(x) = k(-x)$ ,
3.  $\hat{k} \in L_1(\mathbb{R}^1)$ , where  $\hat{k}$  is the Fourier transform of  $k$ ,
4.  $\int_{-\infty}^{\infty} |xk(x)| dx < \infty$ ,
- and
5.  $\int_{-\infty}^{\infty} |xk(x)| dx < 1/2$ .

Then the Fredholm determinant  $D_T(\lambda)$  of the symmetric H-S operator

$$T_\tau f = \int_0^\tau k(x-y)f(y)dy$$

on  $L_2[0, \tau]$  is given by

$$D_\tau(\lambda) = \prod_{j=1}^{\infty} [1 - \lambda \lambda_j(\tau)] = \exp \left[ - \int_0^\tau \alpha(0, s, \lambda) ds \right], \quad (1.1)$$

where  $\alpha(x, \tau, \lambda)$  is the solution of the integral equation:

$$\alpha(x, \tau, \lambda) = \lambda k(x) + \lambda \int_0^\tau k(x-y)\alpha(y, \tau, \lambda) dy.$$

We shall refer to (1.1) as the Kac-Akhiezer formula.

In the above result we observe that the main assumptions on the kernel are:

- (i).  $k(x)$  is a continuous function on  $(-\infty, \infty)$ ;
- (ii). The domain of the integral is an interval in  $\mathbb{R}^1$ ;
- (iii). The kernel is of convolution type.

This result has been extended for the case when  $k(x)$  is only continuous in a neighbourhood of the origin. Further, when  $k(x)$  is not continuous near the origin, a meromorphic function whose poles and zeros correspond to the eigenvalues has been obtained in [5], [7]. Subsequently Rao and Sukavanam [9] have extended these results for normal integral operators. But all these results are with the assumptions (ii) and (iii).

The aim of this paper is to obtain these results by replacing (ii) and (iii) by (ii)' and (iii)' below, where (ii)' is essentially replacing the interval by a suitable smooth domain in  $\mathbb{R}^n$  and (iii)' is essentially replacing the convolution nature of the kernel by an assumption that the map  $x \mapsto k(x, \cdot)$  is continuous.

(ii)':  $\mathcal{C}_1$ : For every  $\tau \geq 0$ ,  $\Omega_\tau$  is a bounded subset of  $\mathbb{R}^n$ .

$\mathcal{C}_2$ :  $\{0\} = \Omega_0 \subseteq \Omega_{\tau_1} \subseteq \Omega_{\tau_2}$  if  $0 \leq \tau_1 \leq \tau_2$ , where  $0$  is the zero vector in  $\mathbb{R}^n$ .

$\mathcal{C}_3$ : The map  $\tau \mapsto \mu(\bar{\Omega}_\tau)$  from the interval  $[0, \infty)$  to  $\mathbb{R}^1$  is absolutely continuous in every bounded interval in  $[0, \infty)$ , where  $\mu$  is the Lebesgue measure on  $\mathbb{R}^n$  and  $\bar{\Omega}_\tau$  is the closure of  $\Omega_\tau$  with respect to the standard topology of  $\mathbb{R}^n$ .

$\mathcal{C}_4$ : For every  $\tau \geq 0$  there exists a surface measure  $\sigma_\tau$  on  $\partial\bar{\Omega}_\tau$ , such that

(i).  $g(t) \stackrel{\text{def}}{=} \int_{\partial\bar{\Omega}_t} f(x) d\sigma_t(x)$  is an integrable function on  $[0, \tau]$  for  $f \in C^0(\bar{\Omega}_\tau)$ , where  $C^0(\bar{\Omega}_\tau)$  is the space of all continuous functions on  $\bar{\Omega}_\tau$ , and

(ii).  $\int_{\bar{\Omega}_\tau} f(x) dx = \int_{t=0}^\tau \left[ \int_{\partial\bar{\Omega}_t} f(x) d\sigma_t(x) \right] dt, \quad \forall f \in C^0(\bar{\Omega}_\tau).$

(A simple motivation for such a family  $\{\Omega_\tau\}_{\tau \geq 0}$  in  $\mathbb{R}^n$  is the ball in  $\mathbb{R}^n$  of radius  $\tau$  and centre at  $0$ .)

(iii)': The kernel is such that

$\mathcal{C}_5$ :  $k(x, y)$  is defined in  $\mathbb{R}^n \times \mathbb{R}^n$ , and is symmetric.

$\mathcal{C}_6$ :  $k$  is locally square integrable.

$\mathcal{C}_7$ : For every  $x \in \mathbb{R}^n$ ,  $k(x, y)$  is well defined and is in  $L_2(\mathbb{R}^n)$ , and for any two compact subsets  $E, F$  of  $\mathbb{R}^n$ , the mapping  $x \mapsto k_x$  from  $E$  into  $L_2(F)$  is continuous; where  $k_x(\cdot) = k(x, \cdot)$ .

By  $\mathcal{C}_5$  and  $\mathcal{C}_6$  it follows that for every  $\tau > 0$ .

$$(K_\tau f)x = \int_{\bar{\Omega}_\tau} k(x, y) f(y) dy \quad (f \in L_2(\bar{\Omega}_\tau))$$

defines a compact self adjoint operator on  $L_2(\bar{\Omega}_\tau)$ . Further we assume

$\mathcal{C}_8$ : For all  $\tau > 0$ , 1 is not in the spectrum of  $K_\tau$ .

Then  $(I - K_\tau)^{-1}$  exists for every  $\tau > 0$  where  $I$  is the identity operator on  $L_2(\bar{\Omega}_\tau)$ .

Let  $\alpha_1(x, y, \tau)$  denote the unique solution of the equation

$$\alpha(x, y, \tau) = k(y, x) + \int_{\bar{\Omega}_\tau} k(y, z) \alpha(x, z, \tau) dz \quad (1.2)$$

and  $\alpha_2(x, y, \tau)$  be that of

$$\alpha(x, y, \tau) = k(y, -x) + \int_{\bar{\Omega}_\tau} k(y, z) \alpha(x, z, \tau) dz. \quad (1.3)$$

Since  $K_\tau$  is a compact self adjoint operator on  $L_2(\bar{\Omega}_\tau)$  it follows from the general theory of such operators that  $K_\tau$  has a discrete set of real eigenvalues  $\{\lambda_i(\tau)\}_{i=1}^\infty$  with the origin as the only possible limit point.

The main results we prove are Theorems 1.1 and 1.2 below:

**Theorem 1.1.** Suppose the domain satisfies conditions  $\mathcal{C}_1 - \mathcal{C}_4$  and the kernel, conditions  $\mathcal{C}_5 - \mathcal{C}_7$ , and suppose that  $\mathcal{C}_8$  holds. Then there exists a function  $u(x, y, \tau)$  such that for each fixed  $x \in \bar{\Omega}_\tau$ ,  $u$  is continuous in the variable  $y$  on  $\bar{\Omega}_\tau$  and

$$u(x, y, \tau) = \alpha_1(x, y, \tau) - k(y, x) \quad (1.4)$$

in the  $L_2$  sense as a function in the variable  $y$  on  $\bar{\Omega}_\tau$  and the corresponding Kac-Akhiezer formula is given by

$$\exp \left\{ - \int_{t=0}^{\tau} dt \int_{\partial \bar{\Omega}_t} u(x, x, t) d\sigma_t(x) \right\} = \prod_{i=1}^{\infty} [\exp(\lambda_i(\tau))] [1 - \lambda_i(\tau)]. \quad (1.5)$$

If  $k$  satisfies the conditions  $\mathcal{C}_1 - \mathcal{C}_7$ , we can apply the Theorem 1.1 to the kernel

$$g_\lambda(x, y) \stackrel{\text{def}}{=} \lambda k(x, y) \quad (\lambda \in \mathbb{R}), \quad (1.6)$$

if  $1/\lambda$  is not in the spectrum of  $K_\tau$ . For the kernel  $g_\lambda$  the eigenvalues will be  $\{\lambda \lambda_n(\tau)\}_{n=1}^\infty$  and the corresponding Kac-Akhiezer formula is given by

$$\begin{aligned} \delta_1(\lambda, \tau) &\stackrel{\text{def}}{=} \exp \left\{ - \int_{t=0}^{\tau} dt \int_{\partial \bar{\Omega}_t} u(x, x, t; \lambda) d\sigma_t(x) \right\} \\ &= \prod_{i=1}^{\infty} [\exp(\lambda \lambda_i(\tau))] [1 - \lambda \lambda_i(\tau)], \end{aligned} \quad (1.7)$$

where  $u(\cdot, \cdot, t; \lambda)$  is as in (1.4) with  $k$  replaced by  $g_\lambda$  and  $\alpha_1$  is the solution of (1.2) with  $k$  replaced by  $g_\lambda$ .

It is to be noted that  $\delta_1(\lambda, \tau)$  has an analytic extension in the complex  $\lambda$ -plane. It follows from (1.7) that  $\lambda_0 \in \mathbb{R}^1 \setminus \{0\}$  is an eigenvalue of  $K_\tau$  if and only if  $1/\lambda_0$  is a zero of this extended function.

Apart from the conditions  $\mathcal{C}_1 - \mathcal{C}_8$  if the  $\Omega_\tau$ 's are symmetric for  $\tau > 0$  and the kernel  $k$  is real valued, we prove

**Theorem 1.2.** Suppose as in Theorem 1.1, conditions  $\mathcal{C}_1 - \mathcal{C}_8$  hold and in addition the domains  $\Omega_\tau$  (for  $\tau > 0$ ) are symmetric and  $k$  is real valued. Then there exists a function  $v(x, y, \tau)$  such that for each fixed  $x \in \bar{\Omega}_\tau$ ,  $v$  is continuous in the variable  $y$  on  $\bar{\Omega}_\tau$  and

$$v(x, y, \tau) = \alpha_2(x, y, \tau) - k(y, -x) \quad (1.8)$$

in the  $L_2$  sense as a function in the variable  $y$  on  $\bar{\Omega}_\tau$  and the corresponding Kac–Akhiezer formula is given by

$$\exp \left\{ - \int_{t=0}^{\tau} dt \int_{\partial \bar{\Omega}_t} v(x, x, t) d\sigma_t(x) \right\} = \frac{\Pi_i^+ [\exp(\lambda_i(\tau))] [1 - \lambda_i(\tau)]}{\Pi_j^- [\exp(\lambda_j(\tau))] [1 - \lambda_j(\tau)]}, \quad (1.9)$$

where  $\Pi_i^+$  and  $\Pi_j^-$  are products over all eigenvalues that have symmetric and anti-symmetric eigenfunctions respectively.

Consider the kernel  $k$  and the family  $\{\bar{\Omega}_\tau\}_{\tau \geq 0}$  as in Theorem 1.2. We can apply Theorem 1.2 to  $g_\lambda(x, y)$ . For this kernel the eigenvalues will be  $\{\lambda \lambda_n(\tau)\}_{n=1}^\infty$  and the corresponding Kac–Akhiezer formula is given by

$$\begin{aligned} \delta_2(\lambda, \tau) &\stackrel{\text{def}}{=} \exp \left\{ - \int_{t=0}^{\tau} dt \int_{\partial \bar{\Omega}_t} v(x, x, t; \lambda) d\sigma_t(x) \right\} \\ &= \frac{\Pi_i^+ [\exp(\lambda \lambda_i(\tau))] [1 - \lambda \lambda_i(\tau)]}{\Pi_j^- [\exp(\lambda \lambda_j(\tau))] [1 - \lambda \lambda_j(\tau)]}, \end{aligned} \quad (1.10)$$

where  $v(\cdot, \cdot, t; \lambda)$  is as in (1.8) with  $k$  replaced by  $g_\lambda$  and  $\alpha_2$  is the solution of (1.3) with  $k$  replaced by  $g_\lambda$ .

It is to be noted that  $\delta_2(\lambda, \tau)$  has a meromorphic extension in the complex  $\lambda$ -plane. It follows from (1.10) that  $\lambda_0 (\neq 0) \in \mathbb{R}^1$  is an eigenvalue of  $K_\tau$  if and only if  $1/\lambda_0$  is a zero or a pole of this function.

We observe that Theorem 1.1 is the generalization of the classical Kac–Akhiezer formula (1.1) as extended by Vittal Rao [8] for kernels continuous near the origin and Theorem 1.2 is the generalization of the meromorphic function extension obtained by Vittal Rao [8].

## 2. Some properties of the eigenvalues

We now obtain some properties of the eigenvalues and eigenfunctions similar to those obtained by Vittal Rao [7]. The ideas behind the proofs are similar to those in [7].

Let  $\lambda_i^+(\tau)$  and  $\lambda_i^-(\tau)$  be respectively the  $i$ th positive and negative eigenvalues of  $K_\tau$ . By the minimax characterization [6], [10] of the eigenvalues we have

**Lemma 2.1.** For each  $i \in \mathbb{N}$ ,  $\lambda_i^+(\tau)$  and  $\lambda_i^-(\tau)$  are respectively nondecreasing and non-increasing functions for  $\tau \geq 0$  if the  $\Omega_\tau$ 's satisfy the conditions  $\mathcal{C}_1 - \mathcal{C}_2$  and the kernel  $k$  satisfies the conditions  $\mathcal{C}_5 - \mathcal{C}_6$ .

**Lemma 2.2.**  $\lambda_i^+(\tau)$  is an absolutely continuous function in the variable  $\tau$  in every bounded

interval of  $[0, \infty)$  for each  $i \in \mathbb{N}$ , if  $k$  is a locally integrable symmetric kernel on  $\mathbb{R}^n \times \mathbb{R}^n$  and the  $\Omega_\tau$ 's satisfy the conditions  $\mathcal{C}_1 - \mathcal{C}_3$ .

*Proof.* For  $\tau \geq 0$  we define

$$F(\tau) = \int_{\bar{\Omega}_\tau} \int_{\bar{\Omega}_\tau} |k(x, y)|^2 dx dy. \quad (2.1)$$

For a  $\tau_0 \in (0, \infty)$  and a finite sequence of disjoint open intervals  $\{(\tau_j, \eta_j)\}_{j=1}^m$  in  $[0, \tau_0]$ , we have

$$\begin{aligned} \sum_{j=1}^m F(\eta_j) - F(\tau_j) &= \sum_{j=1}^m \left\{ \int_{\bar{\Omega}_{\eta_j}} dx \int_{\bar{\Omega}_{\eta_j} \setminus \bar{\Omega}_{\tau_j}} |k(x, y)|^2 dy + \int_{\bar{\Omega}_{\eta_j} \setminus \bar{\Omega}_{\tau_j}} dx \int_{\bar{\Omega}_{\tau_j}} |k(x, y)|^2 dy \right\} \\ &\leq \sum_{j=1}^m \left\{ \int_{\bar{\Omega}_{\tau_0}} dx \int_{\bar{\Omega}_{\eta_j} \setminus \bar{\Omega}_{\tau_j}} |k(x, y)|^2 dy + \int_{\bar{\Omega}_{\eta_j} \setminus \bar{\Omega}_{\tau_j}} dx \int_{\bar{\Omega}_{\tau_0}} |k(x, y)|^2 dy \right\} \\ &= 2 \int_{\bar{\Omega}_{\tau_0}} \int_{\cup_{j=1}^m (\bar{\Omega}_{\eta_j} \setminus \bar{\Omega}_{\tau_j})} |k(x, y)|^2 dx dy \end{aligned} \quad (2.2)$$

and

$$(\mu \times \mu)[\bar{\Omega}_{\tau_0} \times \cup_{j=1}^m (\bar{\Omega}_{\eta_j} \setminus \bar{\Omega}_{\tau_j})] = \mu(\bar{\Omega}_{\tau_0}) \sum_{j=1}^m [\mu(\bar{\Omega}_{\eta_j}) - \mu(\bar{\Omega}_{\tau_j})]. \quad (2.3)$$

Let  $\nu$  be the measure defined on the elementary sets as

$$\nu(E \times F) = \iint_{E \times F} |k(x, y)|^2 dx dy.$$

Since the map  $\tau \mapsto \mu(\bar{\Omega}_{\tau_0})\mu(\bar{\Omega}_\tau)$  is absolutely continuous in  $[0, \tau_0]$  and  $\nu \ll \mu \times \mu$ , using (2.2) and (2.3) one has  $F(\tau)$  is absolutely continuous in the interval  $[0, \tau_0]$ . From (2.1) and the classical theory of H-S operators [6], we have

$$F(\tau) = \sum_{i=1}^{\infty} \lambda_i^2(\tau);$$

hence by Lemma 2.1, we have for  $i \in \mathbb{N}$ , and  $j = 1, 2, 3, \dots, m$ ,

$$F(\eta_j) - F(\tau_j) \geq \lambda_i^2(\eta_j) - \lambda_i^2(\tau_j) \geq 0,$$

from which it follows that  $\lambda_i^2(\tau)$  is an absolutely continuous function in the variable  $\tau$  in the interval  $[0, \tau_0]$ . Hence by Lemma 2.1 we have  $\lambda_i^+(\tau)$  is absolutely continuous in the interval  $[0, \tau_0]$ . By considering  $-K_\tau$  instead of  $K_\tau$  one can conclude the same for the negative eigenvalues. This completes the proof of the lemma. ■

Lemma 2.1 implies that for each  $i \in \mathbb{N}$ ,  $\lambda_i(\tau)$  is differentiable almost everywhere. Let us denote an eigenfunction corresponding to an eigenvalue  $\lambda_i(\tau)$  of  $K_\tau$  by  $\phi_i(\cdot, \tau)$ .

**Lemma 2.3** Let  $i \in \mathbb{N}$  be fixed and  $\lambda_i(\tau)$  be of multiplicity  $m_i(\tau)$ . Then there exist  $m_i(\tau)$  orthonormal eigenfunctions  $\{\phi_{ij}(\cdot, \tau)\}_{j=1}^{m_i(\tau)}$  such that

$$\frac{d\lambda_i(\tau)}{d\tau} = \lambda_i(\tau) \int_{\partial\bar{\Omega}_\tau} |\phi_{ij}(x, \tau)|^2 d\sigma_\tau(x); \quad j = 1, 2, \dots, m_i(\tau), \quad (2.4)$$

for almost every  $\tau$ , if the  $\Omega_\tau$  and  $k$  satisfy the conditions  $\mathcal{C}_1 - \mathcal{C}_7$ .

*Proof.* It is enough if we prove the result for the positive eigenvalues of  $K_\tau$ . Let  $\tilde{\tau} \geq 0$  be such that  $\lambda_i^+(\tau) = 0$  for all  $\tau \leq \tilde{\tau}$  and  $\lambda_i^+(\tau) > 0$  for  $\tau > \tilde{\tau}$ . Hence (2.4) is true for almost every  $\tau \leq \tilde{\tau}$  and for any choice of eigenfunctions. Now  $\lambda_i^+(\tau)$  and  $\mu(\bar{\Omega}_\tau)$  are differentiable almost everywhere by Lemma 2.1 and assumption  $\mathcal{C}_3$  respectively. Let  $\tau_0 > \tilde{\tau}$  be such that both  $d\lambda_i^+(\tau)/d\tau$  and  $d\mu(\bar{\Omega}_\tau)/d\tau$  exist at  $\tau = \tau_0$ . Let  $\{\tau_p\}_{p=1}^\infty$  be an increasing sequence in the interval  $(\tilde{\tau}, \tau_0)$  such that

$$\tau_p \rightarrow \tau_0 \text{ as } p \rightarrow \infty.$$

First let us consider the case when the multiplicity of  $\lambda_i^+(\tau_0)$  namely  $m_i(\tau_0)$  is one. Consider the sequence of eigenvalues  $\{\lambda_i(\tau_p)\}_{p=1}^\infty$  and a corresponding sequence  $\{\phi_i(\cdot, \tau_p)\}_{p=1}^\infty$  of normalized eigenfunctions of  $K_{\tau_p}$ .

We have

$$\lambda_i^+(\tau_p) \phi_i(x, \tau_p) = \int_{\bar{\Omega}_{\tau_p}} k(x, y) \phi_i(y, \tau_p) dy \quad (2.5)$$

and

$$\int_{\bar{\Omega}_{\tau_p}} |\phi_i(x, \tau_p)|^2 dx = 1. \quad (2.6)$$

Though initially the eigenfunctions  $\phi_i(x, \tau_p)$  are defined only in  $\bar{\Omega}_{\tau_p}$ , the right hand side of (2.5) allows us to extend  $\phi_i(x, \tau_p)$  for all  $x \in \mathbb{R}^n$ . Hence without loss of generality we can take each one of the  $\phi_i(\cdot, \tau_p)$  to be defined on  $\bar{\Omega}_{\tau_0}$ .

For each  $x \in \bar{\Omega}_\tau$ , we define a linear functional  $F_{x,\tau}$  on  $L_2(\bar{\Omega}_\tau)$  by

$$F_{x,\tau}(f) = \int_{\bar{\Omega}_\tau} k(x, y) f(y) dy.$$

Condition  $\mathcal{C}_7$  implies that the map  $x \mapsto k_x$  from  $\bar{\Omega}_\tau$  to  $L_2(\bar{\Omega}_\tau)$  is weakly continuous; and hence  $\{k_x\}_{x \in \bar{\Omega}_\tau}$  is a weakly compact subset of  $L_2(\bar{\Omega}_\tau)$ . Hence for  $f \in L_2(\bar{\Omega}_\tau)$ , there exists a constant  $c_{f,\tau} \in (0, \infty)$  such that

$$\sup_{x \in \bar{\Omega}_\tau} \left| \int_{\bar{\Omega}_\tau} k_x(y) f(y) dy \right| \leq c_{f,\tau},$$

which implies that

$$\sup_{x \in \bar{\Omega}} |F_{x,\tau}(f)| \leq c_{f,\tau}.$$

Hence by uniform boundedness principle one can choose a constant  $M_\tau \in (0, \infty)$  such that

$$\|F_{x,\tau}\| \leq M_\tau, \quad \forall x \in \bar{\Omega}_\tau. \quad (2.7)$$



Since

$$\|F_{x,\tau}\|^2 = \int_{\bar{\Omega}_\tau} |k(x,y)|^2 dy,$$

it now follows from (2.7) that,

$\forall \tau > 0$ , there exists a constant  $M_\tau \in (0, \infty)$  such that

$$\int_{\bar{\Omega}_\tau} |k(x,y)|^2 dy \leq M_\tau^2, \quad \forall x \in \bar{\Omega}_\tau. \quad (2.8)$$

We now prove the following properties of the eigenvalue sequence  $\{\phi_i(\cdot, \tau_p)\}_{p=1}^\infty$ :

(a)  $\phi_i(\cdot, \tau_p) \in C^0(\bar{\Omega}_{\tau_0}), \forall p. \quad (2.9)$

(b) There exists a  $J_i \in [0, \infty)$  such that  $|\phi_i(x, \tau_p)| \leq J_i, \forall x \in \bar{\Omega}_{\tau_0}$  and  $\forall p. \quad (2.10)$

(c) There exists a subsequence  $\{\phi_i(\cdot, \tau_q)\}_{q=1}^\infty$  which converges uniformly on  $\bar{\Omega}_{\tau_0}. \quad (2.11)$

*Proof of (a):* For each  $p$ , by  $\mathcal{C}_7$ , the map

$$x \mapsto \int_{\bar{\Omega}_{\tau_p}} k(x,y) \phi_i(y, \tau_p) dy,$$

from the compact set  $\bar{\Omega}_{\tau_0}$  into  $\mathbb{C}$  is continuous; hence by (2.5) the map

$$x \mapsto \lambda_i^+(\tau_p) \phi_i(x, \tau_p)$$

from  $\bar{\Omega}_{\tau_0}$  into  $\mathbb{C}$  is continuous. Now (a) follows as  $\lambda_i^+(\tau_p) > 0$ .

*Proof of (b):* Applying Schwartz inequality to the right hand side of (2.5) and using (2.6) we get

$$|\lambda_i^+(\tau_p) \phi_i(x, \tau_p)|^2 \leq \int_{\bar{\Omega}_{\tau_p}} |k(x,y)|^2 dy \quad \forall x \in \bar{\Omega}_{\tau_0} \text{ and } \forall p. \quad (2.12)$$

Now using  $\mathcal{C}_2$  and (2.12) we get

$$|\lambda_i^+(\tau_p) \phi_i(x, \tau_p)| \leq M_{\tau_0}, \quad \forall x \in \bar{\Omega}_{\tau_0} \text{ and } \forall p$$

where  $M_{\tau_0}$  is as given in (2.8). Since  $\lambda_i^+(\tau_1) > 0$ , (b) follows by Lemma 2.1.

*Proof of (c):* We shall do this in two steps.

*Step 1.* By (b) and Banach-Alaoglu theorem one can find a sub sequence  $\{\phi_i(\cdot, \tau_q)\}_{q=1}^\infty$  of  $\{\phi_i(\cdot, \tau_p)\}_{p=1}^\infty$  and a  $\psi_i \in L_2(\bar{\Omega}_{\tau_0})$  such that  $\phi_i(\cdot, \tau_q) \rightarrow \psi_i$  weakly in  $L_2(\bar{\Omega}_0)$  as  $q \rightarrow \infty$ . Since the operator  $K_{\tau_0}: L_2(\bar{\Omega}_{\tau_0}) \rightarrow L_2(\bar{\Omega}_{\tau_0})$  is compact, it now follows that

$$\|K_{\tau_0} \phi_i(\cdot, \tau_q) - K_{\tau_0} \psi_i\|_{L_2(\bar{\Omega}_{\tau_0})} \rightarrow 0 \text{ as } q \rightarrow \infty.$$

We see that

$$\begin{aligned}
& \left\| \int_{\bar{\Omega}_{\tau_q}} k(\cdot, y) \phi_i(y, \tau_q) dy - \int_{\bar{\Omega}_{\tau_0}} k(\cdot, y) \psi_i(y) dy \right\|_{L_2(\bar{\Omega}_{\tau_0})} \\
& \leq \left\| \int_{\bar{\Omega}_{\tau_0} \setminus \bar{\Omega}_{\tau_q}} k(\cdot, y) \phi_i(y, \tau_q) dy \right\|_{L_2(\bar{\Omega}_{\tau_0})} + \left\| \int_{\bar{\Omega}_{\tau_0}} k(\cdot, y) [\phi_i(y, \tau_q) - \psi_i(y)] dy \right\|_{L_2(\bar{\Omega}_{\tau_0})} \\
& \leq \left( \int_{\bar{\Omega}_{\tau_0}} \int_{\bar{\Omega}_{\tau_0} \setminus \bar{\Omega}_{\tau_q}} |k(x, y)|^2 dx dy \right)^{1/2} [\mu(\bar{\Omega}_{\tau_0} \setminus \bar{\Omega}_{\tau_q})]^{1/2} J_i + \|K_{\tau_0} \phi_i(\cdot, \tau_q) - K_{\tau_0} \psi_i\|_{L_2(\bar{\Omega}_{\tau_0})},
\end{aligned}$$

where  $J_i$  is as given in (2.10). Hence we have

$$\left\| \frac{1}{\lambda_i^+(\tau_q)} \int_{\bar{\Omega}_{\tau_q}} k(\cdot, y) \phi_i(y, \tau_q) dy - \frac{1}{\lambda_i^+(\tau_0)} \int_{\bar{\Omega}_{\tau_0}} k(\cdot, y) \psi_i(y) dy \right\|_{L_2(\bar{\Omega}_{\tau_0})} \rightarrow 0 \text{ as } q \rightarrow \infty;$$

from which it follows that

$$\psi_i = \frac{1}{\lambda_i^+(\tau_0)} K_{\tau_0} \psi_i \quad (2.13)$$

and

$$\|\phi_i(\cdot, \tau_q) - \psi_i\|_{L_2(\bar{\Omega}_{\tau_0})} \rightarrow 0 \text{ as } q \rightarrow \infty. \quad (2.14)$$

From (2.6) it follows that

$$1 = \int_{\bar{\Omega}_{\tau_0}} |\phi_i(x, \tau_q)|^2 dx + \int_{\bar{\Omega}_{\tau_0} \setminus \bar{\Omega}_{\tau_q}} |\phi_i(x, \tau_q)|^2 dx \quad \forall q. \quad (2.15)$$

As

$$\mu(\bar{\Omega}_{\tau_0} \setminus \bar{\Omega}_{\tau_q}) \rightarrow 0 \text{ when } q \rightarrow \infty, \quad (2.16)$$

using (2.10) in (2.15) we get

$$\lim_{q \rightarrow \infty} \int_{\bar{\Omega}_{\tau_0}} |\phi_i(x, \tau_q)|^2 dx = 1. \quad (2.17)$$

Since the map  $f \mapsto \|f\|$  from  $L_2(\bar{\Omega}_{\tau_0})$  to  $\mathbb{R}^1$  is continuous, from (2.14) and (2.17) it follows that

$$\int_{\bar{\Omega}_{\tau_0}} |\psi_i(x)|^2 dx = 1. \quad (2.18)$$

From (2.13) and (2.18) we have  $\psi_i$  is a normalized eigenfunction of  $K_{\tau_0}$  corresponding to the eigenvalue  $\lambda_i^+(\tau_q)$ .

*Step 2.* Let us denote the  $\psi_i$  by  $\phi_i(\cdot, \tau_0)$ . Then (2.13) can be written as

$$\lambda_i^+(\tau_0) \phi_i(x, \tau_0) = \int_{\bar{\Omega}_{\tau_0}} k(x, y) \phi_i(y, \tau_0) dy. \quad (2.19)$$

From (2.5) and (2.19) it follows that for any  $x \in \bar{\Omega}_{\tau_0}$ ,

$$\begin{aligned} & |\lambda_i^+(\tau_q)\phi_i(x, \tau_q) - \lambda_i^+(\tau_0)\phi_i(x, \tau_0)| \\ & \leq \left( \int_{\bar{\Omega}_{\tau_0} \setminus \bar{\Omega}_{\tau_q}} |k(x, y)|^2 dy \right)^{1/2} \left( \int_{\bar{\Omega}_{\tau_0} \setminus \bar{\Omega}_{\tau_q}} |\phi_i(y, \tau_0)|^2 dy \right)^{1/2} \\ & \quad + \left( \int_{\bar{\Omega}_{\tau_q}} |k(x, y)|^2 dy \right)^{1/2} \left( \int_{\bar{\Omega}_{\tau_q}} |\phi_i(y, \tau_q) - \phi_i(y, \tau_0)|^2 dy \right)^{1/2} \\ & \leq M_{\tau_0} J_i [\mu(\bar{\Omega}_{\tau_0} \setminus \bar{\Omega}_{\tau_q})]^{1/2} + M_{\tau_0}^{1/2} \|\phi_i(\cdot, \tau_q) - \phi_i(\cdot, \tau_0)\|_{L_2(\bar{\Omega}_{\tau_0})}. \end{aligned}$$

Now (2.14) and (2.16) imply that

$$\phi_i(x, \tau_q) \rightarrow \phi_i(x, \tau_0) \text{ uniformly on } \bar{\Omega}_{\tau_0} \text{ as } q \rightarrow \infty.$$

This completes the proof of (c).

From (2.5) and ( $\mathcal{G}_4$  ii) it can be easily seen that

$$\begin{aligned} & \frac{\lambda_i^+(\tau_0) - \lambda_i^+(\tau_q)}{\tau_0 - \tau_q} \int_{\bar{\Omega}_{\tau_0}} \phi_i(x, \tau_q) \overline{\phi_i(x, \tau_0)} dx \\ & = \frac{\lambda_i^+(\tau_0)}{\tau_0 - \tau_q} \int_{\bar{\Omega}_{\tau_0} \setminus \bar{\Omega}_{\tau_q}} \{\phi_i(y, \tau_q) - \phi_i(y, \tau_0)\} \overline{\phi_i(y, \tau_0)} dy \\ & \quad + \frac{\lambda_i^+(\tau_0)}{\tau_0 - \tau_q} \int_{\bar{\Omega}_{\tau_0} \setminus \bar{\Omega}_{\tau_q}} |\phi_i(y, \tau_0)|^2 dy \\ & = \frac{\lambda_i^+(\tau_0)}{\tau_0 - \tau_q} \left[ \int_{\tau=\tau_q}^{\tau_0} \int_{\partial \bar{\Omega}_{\tau}} \{\phi_i(y, \tau_q) - \phi_i(y, \tau_0)\} \overline{\phi_i(y, \tau_0)} d\sigma_{\tau}(y) \right. \\ & \quad \left. + \int_{\bar{\Omega}_{\tau_0} \setminus \bar{\Omega}_{\tau_q}} |\phi_i(y, \tau_0)|^2 dy \right]. \end{aligned}$$

Letting  $q \rightarrow \infty$  and using (c) and ( $\mathcal{G}_4$  i), we get (2.4). This proof can be easily modified for the case  $m_i(\tau_0) > 1$  along similar lines as in [7]. ■

### 3. Proofs of the main results

*Proof of Theorem 1.1:* (2.8) implies that the series  $\sum_{i=1}^{\infty} \lambda_i(\tau) \langle \alpha_1, \phi_i \rangle \phi_i(y, \tau)$  converges uniformly and absolutely [6] to  $\int_{\bar{\Omega}} k(y, z) \alpha_1(x, z, \tau) dz$  where

$$\langle \alpha_1, \phi_i \rangle = \int_{\bar{\Omega}} \alpha_1(x, z, \tau) \overline{\phi_i(z, \tau)} dz.$$

Now using (1.2) we get

$$\alpha_1(x, y, \tau) - k(y, x) = \sum_{i=1}^{\infty} \lambda_i(\tau) \langle \alpha_1, \phi_i \rangle \phi_i(y, \tau). \quad (3.1)$$

From the above equation one easily gets

$$\langle \alpha_1, \phi_i \rangle = \frac{\lambda_i(\tau)}{1 - \lambda_i(\tau)} \overline{\phi_i(x, \tau)}. \quad (3.2)$$

Let  $u(x, y, \tau) = \alpha_1(x, y, \tau) - k(y, x)$ . We have from (3.1) and (3.2)

$$u(x, y, \tau) = \sum_{i=1}^{\infty} \frac{(\lambda_i(\tau))^2}{1 - \lambda_i(\tau)} \phi_i(y, \tau) \overline{\phi_i(x, \tau)}. \quad (3.3)$$

For each fixed  $x \in \bar{\Omega}_\tau$ , the series in the right hand side of (3.3) converges uniformly and absolutely in the variable  $y$  and hence the sum is a continuous function in  $y$ . It therefore follows that  $u(x, y, \tau)$  has a continuous representation. Letting  $y \rightarrow x$  in (3.3) and using monotone convergence theorem and Lemma 2.3 we get

$$\int_{\partial \bar{\Omega}_\tau} u(x, x, \tau) d\sigma_\tau(x) = \sum_{i=1}^{\infty} \frac{\lambda_i(\tau)}{1 - \lambda_i(\tau)} \frac{d\lambda_i(\tau)}{d\tau}.$$

By Lemma 2.2 as in [8] we get

$$- \int_{t=0}^{\tau} dt \int_{\partial \bar{\Omega}_t} u(x, x, t) d\sigma_t(x) = \sum_{i=1}^{\infty} \{ \lambda_i(\tau) + \log[1 - \lambda_i(\tau)] \}.$$

Now (1.5) is easily obtained by exponentiating both sides. ■

*Proof of Theorem 1.2:* Since the kernel is real valued, one can choose  $\phi_i$ 's such that they are real and

$$\phi_i(-x, \tau) = \pm \phi_i(x, \tau), \quad \forall x \in \bar{\Omega}_\tau, \quad (3.4)$$

where  $\pm$  correspond respectively to those eigenvalues having symmetric and skew symmetric eigenfunction. By using arguments as in the proof of Theorem 1.1, one can prove that

$$v(x, y, \tau) = \sum_{i=1}^{\infty} \frac{(\lambda_i(\tau))^2}{1 - \lambda_i(\tau)} \phi_i(y, \tau) \overline{\phi_i(-x, \tau)}, \quad (3.5)$$

where  $v(x, y, \tau)$  is a continuous function on  $\bar{\Omega}_\tau$  in the variable  $y$  for each fixed  $x \in \bar{\Omega}_\tau$ . Letting  $y \rightarrow x$  in (3.5) and using (3.4) we get

$$v(x, x, \tau) = \pm \sum_{i=1}^{\infty} \frac{(\lambda_i(\tau))^2}{1 - \lambda_i(\tau)} |\phi_i(x, \tau)|^2.$$

Now by arguments as in the proof of Theorem 1.1 one can arrive at (1.9). ■

#### 4. Some remarks

*Remark 1. (Non symmetric case):* In all the above lemmas if the kernel is not symmetric then the analogous results hold for the singular numbers,  $s_i(\tau)$ . That is, we get for

each  $i = 1, 2, 3, \dots$ ,

- (1).  $s_i(\tau)$  is a nondecreasing function in  $\tau$ ,
- (2).  $s_i(\tau)$  is absolutely continuous in each bounded interval of  $[0, \infty)$ ,
- (3). There exist  $m_i(\tau)$  orthonormal eigenfunctions  $\phi_{ij}(\cdot, \tau) \in L_2(\Omega_\tau)$  corresponding to the eigenvalue  $s_i^2(\tau)$  of  $K_\tau^* K_\tau$  such that

$$\frac{ds_i(\tau)}{d\tau} = \frac{1}{2} s_i(\tau) \int_{\partial\Omega_\tau} |\phi_{ij}(x, \tau)|^2 d\sigma_\tau(x) \quad \text{for } j = 1, 2, 3, \dots, m_i(\tau),$$

and for almost every  $\tau \geq 0$ ; where  $m_i(\tau)$  is the multiplicity of the eigenvalue  $s_i^2(\tau)$ .

*Remark 2.* Lemma 2.3, Theorems 1.1 and 1.2 as well as their proofs remain valid even if we replace the condition  $\mathcal{G}_7$  by a weaker one given below:

Given two compact subsets  $E$  and  $F$  of  $\mathbb{R}^n$  the map  $x \mapsto k_x$  from  $E$  to  $L_2(F)$  is weakly continuous, meaning that,

“for each  $f \in L_2(F)$ , the map  $x \mapsto \int_F k_x(y) \overline{f(y)} dy$  from  $E$  to  $\mathbb{C}$  is continuous.”

### Acknowledgements

The authors are thankful to Dr S Ramaswamy, TIFR Centre, Bangalore, several discussions with whom have helped to sharpen some of the results. The authors thank Prof. M S Ramanujan, Department of Mathematics, University of Michigan, Ann Arbor for his useful comments.

### References

- [1] Akhiezer N I, A continual analogue of some theorems on Toeplitz matrices, *AMS Transl.*, **50** (1966) 295–316
- [2] Dunford N and Schwartz J T, *Linear operators*, Vol II, (Interscience) 1963
- [3] Kac M, Toeplitz-matrices, translation kernels and a related problem in probability theory, *Duke Math. J.*, **21** (1954) 501–509
- [4] Kac M, Theory and application of Toeplitz forms, in *Summer institute on spectral theory and statistical mechanics*, (Brookhaven National Laboratory), (1965) pp. 1–56
- [5] Mullikin T W and Vittal Rao R, Extended Kac-Akhiezer formula for the Fredholm determinant of integral operators, *J. Math. Anal. Appl.*, **61** (1977) 409–415
- [6] Riesz F and Nagy Sz, *Functional analysis*, (Unger, New York) (1955)
- [7] Vittal Rao R, On the eigenvalues of the integral operators with difference kernels, *J. Math. Anal. Appl.* **53** (1976) 554–566
- [8] Vittal Rao R, Extended Akhiezer formula for the Fredholm determinant of difference kernels, *J. Math. Anal. Appl.* **54** (1976) 79–88
- [9] Vittal Rao R and Sukavanam N, Kac-Akhiezer formula for normal integral operators *J. Math. Anal. Appl.*, **114** (1986) 458–467
- [10] Zabraiko P P et al, *Integral equations – a reference text*. (Noordhoff international publishing, Leyden) 1975



# A proof of Howard's conjecture in homogeneous parallel shear flows

MIHIR B BANERJEE, R G SHANDIL and VINAY KANWAR

Department of Mathematics, Himachal Pradesh University, Shimla 171 005, India

MS received 27 September 1993

**Abstract.** A rigorous mathematical proof of Howard's conjecture which states that the growth rate of an arbitrary unstable wave must approach zero as the wave length decreases to zero, in the linear instability of nonviscous homogeneous parallel shear flows, is presented here for the first time under the restriction of the boundedness of the second derivative of the basic velocity field with respect to the vertical coordinate in the concerned flow domain.

**Keywords.** Nonviscous; shear flows.

## 1. Introduction

The point of inflexion theorem of Rayleigh (1880) and the semicircle theorem of Howard (1961) impose necessary restrictions on the basic velocity field  $U(y)$  and the complex wave velocity  $C = C_e + iC_i$ . These are accessible to an arbitrary unstable ( $C_i > 0$ ) wave in the linear instability of nonviscous homogeneous parallel shear flows and it is of interest to have a similar restriction on the growth rate  $kC_i$  possible for such an unstable wave,  $k$  being the wave number and  $y$  being the vertical coordinate. In his pioneering contribution (1961; henceforth referred to as HO), Howard established one such estimate in the form

$$k^2 C_i^2 \leq \max \left( \frac{dU}{dy} \right)^2, \quad (1)$$

and considering its inability to provide the correct qualitative result for plane the Couette flow with  $dU/dy$  constant, which is known to be neutrally stable with  $kC_i \rightarrow 0$  as  $k \rightarrow \infty$ , remarked "This estimate is not usually sharp—for example, the Couette flow with  $dU/dy$  constant, is known to be neutrally stable—but in most cases it will probably give the correct order of magnitude of the maximum growth rate. It is sufficient to show that  $C_i$  must approach zero as wavelength decreases to zero, given the boundedness of  $dU/dy$ ; but there is likelihood that in fact  $kC_i \rightarrow 0$  as  $k \rightarrow \infty$  and... cited in I".

In the present paper we give a rigorous mathematical proof of this conjecture of Howard, namely  $kC_i \rightarrow 0$  as  $k \rightarrow \infty$ , under the restriction of the boundedness of  $d^2U/dy^2$  in the concerned flow domain.

To facilitate reference to HO, we shall make use of the same notation here and denote the basic velocity field by  $U(y)$  and the density field by  $\rho(y)$  while the Rayleigh stability equation that governs the linear instability of nonviscous homogeneous parallel shear flows is (HO; eq. (5.1) with  $\beta = 0$  and  $n = 1$ )

$$\frac{d^2 H}{dy^2} - k^2 H - \frac{d^2 U}{dy^2} H/U - C = 0. \quad (2)$$

The boundary conditions are that  $H$  must vanish on the rigid walls which may recede to  $\pm \infty$  in the limiting cases and thus

$$H(y_1) = H(y_2) = 0. \quad (3)$$

Multiplying (2) by  $H^*$  (the complex conjugate of  $H$ ) throughout and integrating the resulting equation over the vertical range of  $y$  using the boundary conditions (3), we derive

$$\int_{y_1}^{y_2} (|DH|^2 + k^2 |H|^2) dy + \int_{y_1}^{y_2} \frac{d^2 U}{dy^2} |H|^2 dy / U - C = 0. \quad (4)$$

Equating the real part of both sides of (4), we obtain

$$\int_{y_1}^{y_2} (|DH|^2 + k^2 |H|^2) dy + \int_{y_1}^{y_2} \frac{d^2 U}{dy^2} (U - C_e) |H|^2 dy / (U - C_e)^2 + C_i^2 = 0. \quad (5)$$

Now, multiplying (2) by  $d^2 H^* / dy^2$  throughout, we get

$$\frac{d^2 H^*}{dy^2} \left( \frac{d^2 H}{dy^2} - k^2 H \right) - \frac{d^2 H^*}{dy^2} \left( \frac{d^2 U}{dy^2} H/U - C \right) = 0, \quad (6)$$

and substituting for  $d^2 H^* / dy^2$  from (2) in the last term of (6), we derive upon integrating this latter resulting equation over the range of  $y$  with the help of the boundary conditions (3)

$$\begin{aligned} \int_{y_1}^{y_2} (|D^2 H|^2 + k^2 |DH|^2) dy - k^2 \int_{y_1}^{y_2} \frac{d^2 U}{dy^2} |H|^2 dy / U - C \\ - \int_{y_1}^{y_2} \frac{(d^2 U / dy^2)^2 |H|^2 dy}{(U - C_e)^2 + C_i^2} = 0. \end{aligned} \quad (7)$$

Equating the real part of both sides of (7), it follows that

$$\begin{aligned} \int_{y_1}^{y_2} (|D^2 H|^2 + k^2 |H|^2) dy - k^2 \int_{y_1}^{y_2} \frac{(d^2 U / dy^2)^2 (U - C_e) |H|^2 dy}{(U - C_e)^2 + C_i^2} - \\ \int_{y_1}^{y_2} \frac{(d^2 U / dy^2)^2 |H|^2 dy}{(U - C_e)^2 + C_i^2} = 0. \end{aligned} \quad (8)$$



Adding (5) and (8), we arrive at

$$\int_{y_1}^{y_2} (|D^2 H|^2 + 2k^2 |DH|^2 + k^4 |H|^2) dy - \int_{y_1}^{y_2} \frac{(d^2 U/dy^2)^2 |H|^2 dy}{(U - C_y)^2 + C_i^2} = 0, \quad (9)$$

and since for an unstable ( $c_i > 0$ ) wave

$$(U - C_e)^2 + C_i^2 \geq C_i^2, \text{ for all values of } y \text{ such that } y_1 \leq y \leq y_2, \quad (10)$$

it follows from (9) that

$$\int_{y_1}^{y_2} (|D^2 H|^2 + 2k^2 |DH|^2) dy + \int_{y_1}^{y_2} \left[ k^4 - \left( \left( \frac{d^2 U}{dy^2} \right)^2 / C_i^2 \right) \right] |H|^2 dy \leq 0. \quad (11)$$

Equation (11) shows that a necessary condition for its validity is that

$$k^2 C_i \leq \max \left| \frac{d^2 U}{dy^2} \right|, \quad (12)$$

and therefore we must have

$$\lim_{k \rightarrow \infty} k C_i = 0, \quad (13)$$

which proves the validity of the Howard's conjecture.

## References

- [1] Howard L N, Note on a paper of John W. Miles *J. Fluid Mech.* **10** (1961) 509
- [2] Rayleigh L, On the stability or instability of certain fluid motions. *Proc. London Math. Soc.* **11**, (1880) 57



The spectral theory of Schrödinger operators is a highly developed field and has given rise to many new mathematical techniques. Some of these find applications in the scattering theory in quantum mechanics. The scattering theory involves the study of the wave operators and the scattering matrices associated with pairs of self adjoint operators. One of them is often the perturbation of the other by a potential to model the physical scattering of particles. In the last decade there has been a lot of activity in the scattering theory, especially in the understanding of the models of scattering of several particles, analysis of the scattering matrices together with a fine-tuning of the theory for the two particle case. Usually this is done on an Euclidean configuration space. But one can equally well ask similar questions when the configuration space is a Riemannian manifold, in particular a hyperbolic space. In such a case many of the spectral properties of the associated Laplacian (free Hamiltonian) are closely related to the geometry of the manifold and representations of the associated fundamental group. One can also look at the problem from the other end, specifically when the spectral structure is given and one is asked to find the potential in the Schrödinger operator whose spectrum will coincide with the given one. This problem has an analogue in the discretized Schrödinger operator on a lattice. Another interesting problem is the abstract theory of Krein's spectral shift function and the trace formula, which has found many applications in analysis including some results on the index of a pair of operators.

Some of these problems were discussed and a set of solutions were presented in the workshop "Spectral and Inverse Spectral Theory", organized by the Indian Academy of Sciences, Bangalore, during August 24–30, 1993, at the Kodaikanal Observatory of the Indian Institute of Astrophysics. This volume is a result of that effort. Attempts have been made to make this into a kind of review volume, with a few new proofs. It is intended to give an introduction suitable for advanced graduate students and also to serve as a convenient reference for researchers in the area.

We would like to thank the Indian Academy of Sciences for the financial support and encouragement in the concept of a Special Issue, the Institute of Mathematical Sciences, Madras, for help in assembling the participants from different corners of the world, the Indian Institute of Astrophysics for making the pretty and peaceful campus of the Kodaikanal Observatory in the hills of the southern state of Tamil Nadu, available for such an enjoyable meeting. We also wish to put on record our appreciation of the efforts of the contributors in preparing the manuscripts and of the Editor and the staff of the Academy in bringing out this volume in reasonable time.

The year 1994 is also the Diamond Jubilee year of the Academy and it is only appropriate that this volume is christened a Diamond Jubilee special issue of the Proceedings of the Academy (Mathematical Sciences).

Kalyan B Sinha  
M Krishna  
(Guest Editors)



## Scattering theory for Stark Hamiltonians

ARNE JENSEN

Institut Mittag-Leffler, Auravägen 17, S-182 62 Djursholm, Sweden

**Abstract.** We give an introduction to the spectral and scattering theory for Schrödinger operators. An abstract short range scattering theory is developed. It is applied to perturbations of the Laplacian. Particular attention is paid to the study of Stark Hamiltonians. The main result is an explanation of the discrepancy between the classical and the quantum scattering theory for one-dimensional Stark Hamiltonians.

**Keywords.** Spectral theory; scattering theory; Schrödinger operators.

### 1. Introduction

In this paper we give a short survey of potential scattering theory in nonrelativistic quantum mechanics. Particular attention is paid to Stark Hamiltonians. The reason for this is a remarkable discrepancy between classical and quantum scattering in the one-dimensional case. In the last section this discrepancy is resolved.

Let  $H$  be a selfadjoint operator, called the Hamiltonian, on a Hilbert space  $\mathcal{H}$ . The time-dependent Schrödinger equation

$$i \frac{d}{dt} \varphi(t) = H \varphi(t), \quad (1.1)$$

$$\varphi(0) = \varphi_0, \quad (1.2)$$

is solved by

$$\varphi(t) = U(t) \varphi_0 = e^{-itH} \varphi_0. \quad (1.3)$$

In potential scattering theory we have a free Hamiltonian  $H_0$  and a potential  $V$  such that the full Hamiltonian is  $H = H_0 + V$ . The free evolution is given by  $U_0(t) = \exp(-itH_0)$ . In scattering theory we compare the two evolutions  $U_0(t)$  and  $U(t)$  for large positive and negative times  $t$ , using the wave operators

$$W_{\pm} = s\text{-}\lim_{t \rightarrow \pm \infty} U(-t) U_0(t) P_{ac}(H_0) = s\text{-}\lim_{t \rightarrow \pm \infty} e^{itH} e^{-itH_0} P_{ac}(H_0). \quad (1.4)$$

Here  $P_{ac}(H_0)$  denotes the projection onto the subspace of absolute continuity of  $H_0$ . An important property of the wave operators is the intertwining relation

$$e^{itH} W_{\pm} = W_{\pm} e^{itH_0}. \quad (1.5)$$

Assume that the wave operators exist. They are said to be complete, if

$\text{Ran } W_{\pm} = \mathcal{H}_{\text{ac}}(H)$ , the subspace of absolute continuity of  $H$ . In that case the absolutely continuous parts of  $H_0$  and  $H$  are unitarily equivalent via the wave operators.

Let us briefly describe the contents of the paper. In §2 we give some definitions and basic results from spectral theory. In §3 we give general properties of the wave operators. The results are well-known, however to make the paper self-contained we give some of the short proofs.

In §4 we continue the study of scattering in the abstract setting. We develop an abstract short range scattering theory based on propagation estimates for  $H_0$  relative to a conjugate operator  $A$ . The short range condition on  $V$  is then a condition relative to  $A$ . This concept of short range potential is more general than the one used in concrete settings, for example some oscillating slowly decaying potentials are short range perturbations of the Laplacian in our theory. The abstract theory seems to have several advantages. The proofs are short and straightforward generalizations of the methods introduced by Enss [11] and Mourre [32]. Furthermore, the assumptions are fairly easy to verify in many important cases, and the theory covers many interesting applications. The main theorem in this abstract theory was stated without proof in [23]. A short presentation was given in [22].

Sections 5–10 are devoted to applications of the abstract theory, and to further developments of scattering theory. We start in §5 by showing that the usual short range scattering theory for  $H_0 = -\frac{1}{2}\Delta$  and  $V$  multiplication by a real valued function on  $\mathcal{H} = L^2(\mathbf{R}^d)$  is an easy consequence of the abstract theory.

In §6 we begin our discussion of Stark Hamiltonians. The free Hamiltonian is

$$H_0 = -\frac{1}{2}\Delta + F \cdot x, \quad F \in \mathbf{R}^d \setminus \{0\}, \quad (1.6)$$

and for a real-valued function  $\Phi$  the full Hamiltonian is

$$H = -\frac{1}{2}\Delta + F \cdot x + \Phi(x). \quad (1.7)$$

In §6.1 the usual short range scattering theory for Stark Hamiltonians is derived from our abstract theory. In rough asymptotic form the short range conditions are

$$\Phi(x_{\parallel}, x_{\perp}) = \begin{cases} O(|x_{\parallel}|^{-1/2-\varepsilon}) & \text{as } x_{\parallel} \rightarrow -\infty, \\ o(|x_{\parallel}|) & \text{as } x_{\parallel} \rightarrow \infty, \\ o(1) & \text{as } |x_{\perp}| \rightarrow \infty. \end{cases} \quad (1.8)$$

where we decompose  $x = (x_{\parallel}, x_{\perp})$  into a component  $x_{\parallel}$  in the direction of  $F$  and one orthogonal to  $F$ . This class of potentials is the one considered in [4, 47].

In §6.2 we look in detail at the one-dimensional case. If  $\Phi(x_1) = W''(x_1)$ , where  $W$  is a bounded function with four bounded derivatives, then we show that the wave operators exist and are unitary. This class of potentials includes periodic and some almost-periodic functions. This result was first obtained in [19]. Further results can be found in [20, 43], including some results on oscillating potentials in dimensions  $d \geq 2$ .

In §6.3 we then look at the corresponding classical scattering problem with the classical Hamiltonian  $H(x_1, p_1) = \frac{1}{2}p_1^2 + x_1 + \Phi(x_1)$ . For potentials  $\Phi = W''$  there is complete analogy, since existence and completeness of the classical wave operators also holds for this class. For decaying potentials the situation is different. We prove

that under the condition

$$|\Phi(x_1)| + |\Phi'(x_1)| \leq C(\log(2 + |x_1|))^{-\beta}, \quad \beta > 1, \quad (1.9)$$

the classical wave operators exist and are complete. This result was first obtained in [25]. Thus there is a considerable discrepancy between the decay rate in the quantum case (1.8) and the classical case (1.9). The quantum result cannot be improved for potentials without oscillation, since by the result in [34] the quantum wave operators do not exist for homogeneous potentials  $\Phi(x) \sim \lambda|x|^{-\gamma}$ ,  $\lambda \neq 0$ ,  $0 < \gamma \leq 1/2$ . This result holds in any dimension. Generalizations can be found in [25, 24].

In order to investigate this discrepancy in detail we introduce the moving frame picture for Stark Hamiltonians in § 7. The scattering theory for the pair  $(H, H_0)$  given by (1.7) and (1.6) is equivalent to the scattering theory for the pair  $(K(t), K_0)$ , where  $K_0 = -\frac{1}{2}\Delta$  and  $K(t) = K_0 + V(t, x)$  with  $V(t, x) = \Phi(x - \frac{1}{2}t^2 F)$ . Thus in § 8 we study existence and non-existence of the wave operators for the pair  $(K(t), K_0)$  for a general class of time-dependent potentials  $V(t, x)$ . An application of the non-existence result yields the non-existence of wave operators for potentials  $\Phi(x) \sim \lambda|x|^{-\gamma}$  mentioned above. Another application recovers the well-known result that the Coulomb potential  $c/|x|$  is the borderline case for perturbations of  $-\Delta$  without oscillation. For such long range potentials Dollard proposed a modified free evolution. In § 9 we study Dollard-type modified wave operators for time-dependent long range potentials, assuming a decomposition  $V(t, x) = V_s(t, x) + V_l(t, x)$ . The modified free evolution is given by

$$U_D(t) = \exp\left(-it\frac{1}{2}p^2 - i\int_0^t V_l(\tau, \tau p) d\tau\right). \quad (1.10)$$

Here  $p = -i\nabla$  as usual.

Finally in § 10 we put all the pieces together. Graf [12] proposed to explain the discrepancy by using the following decomposition of  $V(t, x) = \Phi(x - \frac{1}{2}t^2 F)$  in the moving frame picture:

$$V_l(t, x) = \Phi(-\frac{1}{2}t^2 F), \quad (1.11)$$

$$V_s(t, x) = \Phi(x - \frac{1}{2}t^2 F) - \Phi(-\frac{1}{2}t^2 F). \quad (1.12)$$

We write

$$S_G(t) = -\int_0^t \Phi(-\frac{1}{2}\tau^2 F) d\tau$$

for the modifier thus obtained. Graf proves existence and completeness of the wave operators

$$W_{\pm}^G = s\text{-}\lim_{t \rightarrow \pm\infty} e^{itH} e^{-itH_0 + iS_G(t)}$$

under the assumption  $\nabla\Phi(x) = O(|x|^{-1-\varepsilon})$  as  $|x| \rightarrow \infty$ .

The main result in § 10 shows that in the one-dimensional case one can prove existence of  $W_{\pm}^G$  under the condition (1.9). This resolves the discrepancy, since the

pure phase correction  $\exp(iS_G(t))$  is not observable. These results and further developments were first obtained in [24].

If one assumes the stronger conditions that  $\Phi$  is smooth and for all multi-indices  $\alpha$

$$|\partial_x^\alpha \Phi(x_\parallel, x_\perp)| \leq C_\alpha (1 + |x_\parallel|)^{-|\alpha|/2 - \varepsilon} \quad (1.13)$$

and  $\Phi(x) \rightarrow 0$  as  $|x| \rightarrow \infty$ , then in the one-dimensional case asymptotic completeness also holds, a result which is obtained by combining the results here with the result in [26]. Under these stronger conditions White [43, 44] proved existence and completeness of the Dollard-type modified wave operators in any dimension, however with a nontrivial modification in the  $x_\perp$  variable. One may ask whether the Graf modification can be used in this case. In § 10 we show that this is not possible, by proving a non-existence theorem for Graf-modified wave operators.

Finally let us give some references to comprehensive treatments of scattering theory. Monographs on scattering theory include [2, 3, 5, 36, 38, 45]. The paper [41] contains a large number of applications of the Enss method. There are also results on scattering theory in the monograph [8].

Recently there has been significant developments on the scattering problem for the  $N$ -body problem. It is beyond the scope of this paper to discuss these.

## 2. Notation and preliminary results

In this section we give some remarks on the background needed to read this paper, and introduce notation which will be used throughout the paper.

As a general background we assume familiarity with functional analysis at the level of the book [39]. We also need some more specific results on perturbation theory. The results in §§ 2 and 3 in [37, Chapter X] will suffice. For other results needed we will give explicit references.

Let  $H$  be a selfadjoint operator on the separable Hilbert space  $\mathcal{H}$  with domain  $\mathcal{D}(H)$ . We denote the resolvent by  $R(z) = (H - z)^{-1}$ , the resolvent set by  $\rho(H)$ , and the spectrum by  $\sigma(H)$ . The spectral measure of  $H$  is denoted  $E$ , or sometimes  $E_H$ . The spectral theorem gives the representation.

$$H = \int \lambda dE(\lambda),$$

and the functional calculus is given by

$$f(H) = \int f(\lambda) dE(\lambda)$$

for all Borel functions  $f$  on  $\mathbf{R}$ .

### 2.1 Parts of the spectrum

The usual decomposition of a measure relative to the Lebesgue measure on  $\mathbf{R}$  gives rise to the notions of singular spectrum and absolutely continuous spectrum. Since



we need these results, we will state them in some detail. The Lebesgue measure on  $\mathbf{R}$  is denoted by  $|\cdot|$ . We define the *absolutely continuous* subspace by

$$\mathcal{H}_{ac}(H) = \{\psi \in \mathcal{H} \mid (\psi, E_H(\cdot)\psi) \text{ is absolutely continuous w.r.t. } |\cdot|\}. \quad (2.1)$$

The *singular* subspace is defined by

$$\mathcal{H}_s(H) = \{\psi \in \mathcal{H} \mid (\psi, E_H(\cdot)\psi) \text{ is singular w.r.t. } |\cdot|\}. \quad (2.2)$$

The point-spectral subspace is given by

$$\mathcal{H}_p(H) = \text{Closed subspace spanned by all eigenvectors of } H. \quad (2.3)$$

It is clearly a subspace of  $\mathcal{H}_s(H)$ . The singular continuous subspace is obtained by taking orthogonal complement.

$$\mathcal{H}_{sc}(H) = \mathcal{H}_s(H) \ominus \mathcal{H}_p(H). \quad (2.4)$$

The continuous subspace is given by

$$\mathcal{H}_c(H) = \mathcal{H} \ominus \mathcal{H}_p(H). \quad (2.5)$$

Let  $\mathcal{K}$  be a closed subspace of  $\mathcal{H}$ . Its orthogonal complement is denoted  $\mathcal{K}^\perp$ . The subspace  $\mathcal{K}$  is said to be *invariant* under  $H$ , if  $H(\mathcal{K} \cap \mathcal{D}(H)) \subseteq \mathcal{K}$ . A subspace  $\mathcal{K}$  is said to be *reducing* for  $H$ , if both  $\mathcal{K}$  and  $\mathcal{K}^\perp$  are invariant under  $H$ , and, furthermore, we have

$$\mathcal{D}(H) = (\mathcal{K} \cap \mathcal{D}(H)) \oplus (\mathcal{K}^\perp \cap \mathcal{D}(H)).$$

**Lemma 2.1.** *Let  $P$  denote the orthogonal projection onto  $\mathcal{K}$ . Then  $\mathcal{K}$  is reducing if and only if  $TP \supset PT$ .*

*Proof.* See for example [27].

If  $\mathcal{K}$  is a reducing subspace for  $H$ , then the restriction of  $H$  to  $\mathcal{K}$  is denoted  $H_{\mathcal{K}}$  (as an operator in the Hilbert space  $\mathcal{K}$ ).

## PROPOSITION 2.2

*Let  $\mathcal{K}$  be a reducing subspace for the selfadjoint operator  $H$ . Then  $H_{\mathcal{K}}$  and  $H_{\mathcal{K}^\perp}$  are selfadjoint. Furthermore,  $\sigma(H) = \sigma(H_{\mathcal{K}}) \cup \sigma(H_{\mathcal{K}^\perp})$ . The subspace  $\mathcal{K}$  is reducing if and only if  $E_H(I)P = PE_H(I)$  for all intervals  $I \subseteq \mathbf{R}$ .*

*Proof.* See for example [27].

With these results we can now state a theorem:

**Theorem 2.3.**  $\mathcal{H}_{ac}(H)$  and  $\mathcal{H}_s(H)$  are closed subspaces of  $\mathcal{H}$ , are orthogonal complements to each other, and both reduce  $H$ .

**Remark 2.4.** Clearly  $\mathcal{H}_p(H)$  reduces  $H$ , hence both  $\mathcal{H}_c(H)$  and  $\mathcal{H}_{sc}(H)$  reduce  $H$ .

*Proof.* See for example [27].

Let us record the relations between these spaces:

$$\mathcal{H} = \mathcal{H}_{ac}(H) \oplus \mathcal{H}_{sc}(H) \oplus \mathcal{H}_p(H) \quad (2.6)$$

$$= \mathcal{H}_{ac}(H) \oplus \mathcal{H}_s(H) \quad (2.7)$$

$$= \mathcal{H}_c(H) \oplus \mathcal{H}_p(H). \quad (2.8)$$

The operators obtained by restriction to these spaces are denoted

$$H_{ac}, \quad H_{sc}, \quad H_p, \quad H_s, \quad H_c.$$

In four of the five cases we use these spaces to define components of the spectrum of  $H$ .

$$\sigma_{ac}(H) = \sigma(H_{ac}), \quad \text{the absolutely continuous spectrum,} \quad (2.9)$$

$$\sigma_{sc}(H) = \sigma(H_{sc}), \quad \text{the singular continuous spectrum,} \quad (2.10)$$

$$\sigma_c(H) = \sigma(H_c), \quad \text{the continuous spectrum,} \quad (2.11)$$

$$\sigma_s(H) = \sigma(H_s), \quad \text{the singular spectrum.} \quad (2.12)$$

One would perhaps expect that the point spectrum would be defined by  $\sigma_p(H) = \sigma(H_p)$ . However, this is not the case, and the usual definition (which we follow here) is

$$\sigma_{pp}(H) = \text{All eigenvalues of } H. \quad (2.13)$$

It is called the *pure point spectrum*. The reason for this convention is that the spectrum of an operator is a closed set, but in some interesting cases the eigenvalues form a dense subset of (part of) the spectrum of  $H$ . By the definitions  $\sigma_{ac}(H)$ ,  $\sigma_{sc}(H)$ ,  $\sigma_c(H)$ , and  $\sigma_s(H)$  are closed subsets of  $\mathbf{R}$ . Simple examples using direct sums show that the relative position of these parts of the spectrum within  $\sigma(H)$  can be arbitrary. We record the following relations:

$$\sigma(H) = \sigma_{ac}(H) \cup \sigma_s(H). \quad (2.14)$$

$$\sigma(H) = \sigma_{ac}(H) \cup \sigma_{sc}(H) \cup \overline{\sigma_{pp}(H)}. \quad (2.15)$$

$$\sigma(H) = \sigma_c(H) \cup \overline{\sigma_{pp}(H)}. \quad (2.16)$$

There is another decomposition of the spectrum which is important. We have the following definition:

## DEFINITION 2.5

Let  $H$  be a selfadjoint operator on  $\mathcal{H}$ . The discrete spectrum  $\sigma_d(H)$  consists of all isolated eigenvalues of finite multiplicity for  $H$ . The essential spectrum is defined by  $\sigma_{ess}(H) = \sigma(H) \setminus \sigma_d(H)$ .

*Remark 2.6.* One should note that only in the case of selfadjoint operators is the above definition of essential spectrum the “good” one. This is due to the fact that considered as a subset of the complex plane the boundary of the set  $\sigma(H)$  is equal to set itself. For general closed operators a good definition of essential spectrum can be found in [27]. See also the discussion in [13].

Let  $H$  be a selfadjoint operator on  $\mathcal{H}$ . Then we have

(i)  $\sigma_{\text{ess}}(H)$  is a closed subset of  $\mathbf{R}$ .

(ii)  $\lambda \in \sigma_{\text{ess}}(H)$  if and only if there exists an orthonormal sequence  $\{\varphi_n\}$ ,  $\varphi_n \in \mathcal{D}(H)$ , such that  $\|(H - \lambda)\varphi_n\| \rightarrow 0$  as  $n \rightarrow \infty$ .

*Proof.* See for example [39].

The following result is a version of Weyl's theorem on the stability of the essential spectrum.

**Theorem 2.8.** Let  $H_j$ ,  $j = 1, 2$ , be selfadjoint operators on  $\mathcal{H}$ . Assume there exists  $z \in \mathbf{C}$ ,  $\text{Im } z \neq 0$ , such that  $(H_2 - z)^{-1} - (H_1 - z)^{-1}$  is a compact operator on  $\mathcal{H}$ . Then  $\sigma_{\text{ess}}(H_1) = \sigma_{\text{ess}}(H_2)$ .

*Proof.* See for example [39].

## 2.2 Partial isometries

Partial isometries play an important role in our discussions. We record a definition and a few properties for future reference. The bounded operators on a Hilbert space  $\mathcal{H}$  are denoted  $\mathcal{B}(\mathcal{H})$ . Below we give the definition in the single Hilbert space case. There is an obvious generalization to an operator between two Hilbert spaces of both the definition and the proposition below.

### DEFINITION 2.9

$W \in \mathcal{B}(\mathcal{H})$  is called a partial isometry, if there exists a closed subspace  $\mathcal{M}$  of  $\mathcal{H}$  such that  $\|W\psi\| = \|\psi\|$  for all  $\psi \in \mathcal{M}$  and  $W\psi = 0$  for all  $\psi \in \mathcal{M}^\perp$ .

Note that the range  $\text{Ran}(W)$  is a closed subspace of  $\mathcal{H}$ . It is called the *final subspace* of  $W$ . Clearly  $\mathcal{M} = \text{Ker}(W)^\perp$ . It is called the *initial subspace* of  $W$ .

### PROPOSITION 2.10

Let  $W \in \mathcal{B}(\mathcal{H})$ . The following statements are equivalent.

- (i)  $W$  is a partial isometry with initial subspace  $\mathcal{M}$  and final subspace  $\mathcal{N}$ .
- (ii)  $\text{Ran}(W) = \mathcal{N}$  and  $W^*W = P_{\mathcal{M}}$ . ( $P_{\mathcal{M}}$  denotes the orthogonal projection onto  $\mathcal{M}$ .)
- (iii)  $W^*W = P_{\mathcal{M}}$  and  $WW^* = P_{\mathcal{N}}$ .
- (iv)  $W^*$  is a partial isometry with initial subspace  $\mathcal{N}$  and final subspace  $\mathcal{M}$ .

*Proof.* See for example [39].

## 3. Scattering theory: General results

In this section we give general results on scattering theory. It is possible to discuss scattering theory for a large number of different systems from both classical mechanics

and quantum mechanics. In this section we will restrict ourselves to the discussion of the scattering theory for two unitary evolution groups in one Hilbert space. This framework is intended for the study of two-body Schrödinger operators. Classical scattering theory will be discussed in § 6.3.

The results in this section are part of the general abstract scattering theory and can be found in any of the textbooks on scattering theory, for example [2, 3, 5, 38, 45]. We mainly follow [27] in our presentation of the results.

### 3.1 The wave operators: General results

We start with a selfadjoint operator  $H_1$  on a separable Hilbert space  $\mathcal{H}$ . The Schrödinger equation is the evolution problem

$$i \frac{d}{dt} \varphi(t) = H_1 \varphi(t), \quad (3.1a)$$

$$\varphi(0) = \varphi_0. \quad (3.1b)$$

By Stone's theorem the solution is given as

$$\varphi(t) = e^{-itH_1} \varphi_0 \quad (3.2)$$

if and only if  $\varphi_0 \in \mathcal{D}(H_1)$ . In many cases it is more convenient to work with the unitary group  $U_1(t) = \exp(-itH_1)$ , since it is defined on all of the Hilbert space. In scattering theory we consider the problem (3.1) for two selfadjoint operators  $H_j$ ,  $j = 1, 2$ . The associated unitary groups are denoted  $U_j(t) = \exp(-itH_j)$ . We want to compare the solutions  $\varphi(t) = U_1(t)\varphi_0$  and  $\psi(t) = U_2(t)\psi_0$  as  $t \rightarrow \pm \infty$ . Usually we require norm convergence in the Hilbert space, such that we look at the question whether

$$\|\varphi(t) - \psi(t)\| = \|U_1(t)\varphi_0 - U_2(t)\psi_0\| \rightarrow 0 \quad \text{as } t \rightarrow \pm \infty$$

This leads us to the operator

$$W(t) = W(H_2, H_1; t) = U_2(-t)U_1(t) \quad (3.3)$$

and to the question whether the limits

$$W_{\pm} = W_{\pm}(H_2, H_1) = s\text{-}\lim_{t \rightarrow \pm \infty} W(H_2, H_1; t) \quad (3.4)$$

exist. These operators are called the wave operators for the pair  $(H_2, H_1)$ . They are the main object of investigation in this paper.

We start by motivating a slight change in the definition (3.4). In general, the limit (3.4) can only exist, if  $H_1$  has purely absolutely continuous spectrum. This is partly justified by the following argument: Assume  $H_1 \varphi = \lambda \varphi$ ,  $\varphi \neq 0$ . Then  $\exp(-it(H_1 - \lambda))\varphi = \varphi$ . Assume that  $W_+$  from (3.4) exists. In that case for all  $s \in \mathbf{R}$

$$\|W(t+s)\varphi - W(t)\varphi\| = \|e^{is(H_2 - \lambda)}\varphi - \varphi\| \rightarrow 0$$

as  $t \rightarrow \infty$ , hence  $e^{is(H_2 - \lambda)}\varphi = \varphi$  for all  $s$ . Stone's theorem implies  $\varphi \in \mathcal{D}(H_2)$  and  $H_2 \varphi = \lambda \varphi$ . We conclude that all eigenvalues of  $H_1$  must also be eigenvalues of  $H_2$  with the same eigenfunctions.

For this reason the projection onto the absolutely continuous subspace is introduced into the definition. Let  $P_j$  denote the orthogonal projection onto the space  $\mathcal{H}_{ac}(H_j)$ ,  $j = 1, 2$ . The choice of  $\mathcal{H}_{ac}(H_j)$  instead of  $\mathcal{H}_c(H_j)$  is partly a matter of mathematical convenience. In many cases we are interested in operators with empty singular continuous spectrum. For further discussion of this point, and for geometric characterization of scattering subspaces, we refer to [1, 13].

Thus we are led to the definition

### DEFINITION 3.1

$$W_{\pm} = W_{\pm}(H_2, H_1) = s\text{-}\lim_{t \rightarrow \pm \infty} W(H_2, H_1; t) P_1. \quad (3.5)$$

In many cases, for example  $H_1 = -\Delta$  on  $\mathcal{H} = L^2(\mathbf{R}^d)$ , we have  $P_1 = 1$  (the identity operator), hence there is no difference between (3.4) and (3.5).

We state the main properties of the wave operators in this general framework. We state the theorem only for  $W_+$ . Entirely analogous results hold for  $W_-$ .

**Theorem 3.2.** *Assume that  $W_+$  exists. Then  $W_+$  is a partial isometry with initial subspace  $\mathcal{H}_{ac}(H_1)$  and final subspace  $\mathcal{N}_+$  contained in  $\mathcal{H}_{ac}(H_2)$ . The subspace  $\mathcal{N}_+$  reduces  $H_2$ . Let  $Q_+$  denote the orthogonal projection onto  $\mathcal{N}_+$ . Then we have*

$$W_+^* W_+ = P_1, \quad W_+ W_+^* = Q_+, \quad (3.6)$$

$$W_+ = W_+ P_1 = Q_+ W_+ = P_2 W_+, \quad (3.7)$$

$$W_+^* = P_1 W_+^* = W_+^* Q_+ = W_+^* P_2. \quad (3.8)$$

Furthermore, we have

$$H_2 W_+ \supseteq W_+ H_1, \quad H_1 W_+^* \supseteq W_+^* H_2. \quad (3.9)$$

In particular,  $H_{1,ac}$  is unitarily equivalent via  $W_+$  to the part of  $H_2$  in  $\mathcal{N}_+$ , and  $\sigma_{ac}(H_1) \subseteq \sigma_{ac}(H_2)$ .

**Remark 3.3.** We use the notation  $H_{j,ac}$  for the restriction of  $H_j$  to  $\mathcal{H}_{ac}(H_j)$ ,  $j = 1, 2$ .

*Proof.* We have  $\|W_+ \varphi\| = \|P_1 \varphi\|$  for all  $\varphi \in \mathcal{H}$ , hence  $W_+$  is a partial isometry with initial subspace  $P_1 \mathcal{H}$ , and  $W_+^* W_+ = P_1$ . Let  $\mathcal{N}_+ = \text{Ran}(W_+)$ . Then  $Q_+ = W_+ W_+^*$  (see Proposition 2.10).

For any  $s, t \in \mathbf{R}$

$$e^{isH_2} W(t) = W(t+s) e^{isH_1},$$

so taking strong limit we get

$$e^{isH_2} W_+ = W_+ e^{isH_1}, \quad s \in \mathbf{R}. \quad (3.10)$$

Using the Laplace transform representation of the resolvent we get

$$(H_2 - z)^{-1} W_+ = W_+ (H_1 - z)^{-1}, \quad \text{Im } z \neq 0. \quad (3.11)$$

$$H_2 W_+ \supseteq W_+ H_1.$$

Taking adjoints we get

$$H_1 W_+^* \supseteq W_+^* H_2.$$

Using these results we have

$$Q_+ H_2 = W_+ W_+^* H_2 \subseteq W_+ H_1 W_+^* \subseteq H_2 W_+ W_+^* = H_2 Q_+.$$

By Lemma 2.1  $\mathcal{N}_+$  reduces  $H_2$ .

Now we have  $Q_+ H_2 Q_+ \subseteq H_2 Q_+ Q_+ = H_2 Q_+$ , but since these operators have the same domain, we must have  $Q_+ H_2 Q_+ = H_2 Q_+$ . Using this result and the above computations, we conclude

$$Q_+ H_2 Q_+ = W_+ H_1 W_+^*$$

and therefore

$$\begin{aligned} H_2 W_+ &= H_2 Q_+ W_+ = Q_+ H_2 Q_+ W_+ \\ &= W_+ H_1 W_+^* W_+ = W_+ H_1 P_1. \end{aligned}$$

Let  $E_j$  denote the spectral measure of  $H_j$ . Equation (3.11) together with Stone's formula for the spectral measure and standard measure theory arguments gives

$$E_2(B) W_+ = W_+ E_1(B), \quad B \subseteq \mathbf{R} \quad \text{Borel set.} \quad (3.12)$$

Thus for all  $\varphi \in \mathcal{H}$

$$\|E_2(B) W_+ \varphi\| = \|W_+ E_1(B) \varphi\| = \|P_1 E_1(B) \varphi\| = \|E_1(B) P_1 \varphi\|$$

and we conclude  $\mathcal{N}_+ \subseteq \mathcal{H}_{\text{ac}}(H_2)$ . This completes the proof, since now we can use  $P_2 Q_+ = Q_+$  together with the above computations.

*Remark 3.4.* The relations (3.9)–(3.12) are all called the *intertwining relation* or *intertwining property*, as is the following equation

$$f(H_2) W_+ = W_+ f(H_1), \quad (3.13)$$

which holds for all bounded Borel functions on  $\mathbf{R}$ .

### DEFINITION 3.5

Assume  $W_+$  exists. Then  $W_+$  is said to be complete, if  $\mathcal{N}_+ = \text{Ran}(W_+) = \mathcal{H}_{\text{ac}}(H_2)$ .  $W_+$  is said to be strongly complete, if furthermore  $\sigma_{\text{sc}}(H_2) = \emptyset$ .

These results and the definition motivate why we are interested in proving completeness of the wave operators. If completeness holds, the absolutely continuous parts of  $H_2$  and  $H_1$  are unitarily equivalent, the equivalence being obtained via  $W_+$ .

There are some properties of the wave operators which we will need. We state the results for  $W_+$ . Entirely analogous results hold for  $W_-$ , with the obvious modification

of replacing  $t \rightarrow +\infty$  by  $t \rightarrow -\infty$ . The following result is known as the *chain rule* for wave operators. Here we have three selfadjoint operators and use the complete notation for the wave operators.

**Theorem 3.6.** *If  $W_+(H_2, H_1)$  and  $W_+(H_3, H_2)$  exist, then  $W_+(H_3, H_1)$  exists, and*

$$W_+(H_3, H_1) = W_+(H_3, H_2) W_+(H_2, H_1). \quad (3.14)$$

*Proof.* By definition

$$W_+(H_2, H_1) = s\text{-}\lim_{t \rightarrow \infty} e^{itH_2} e^{-itH_1} P_1$$

$$W_+(H_3, H_2) = s\text{-}\lim_{t \rightarrow \infty} e^{itH_3} e^{-itH_2} P_2.$$

Hence

$$\begin{aligned} W_+(H_3, H_2) W_+(H_2, H_1) &= s\text{-}\lim_{t \rightarrow \infty} e^{itH_3} e^{-itH_2} P_2 e^{itH_2} e^{-itH_1} P_1 \\ &= s\text{-}\lim_{t \rightarrow \infty} e^{itH_3} P_2 e^{-itH_1} P_1. \end{aligned}$$

It remains to prove

$$s\text{-}\lim_{t \rightarrow \infty} (1 - P_2) e^{-itH_1} P_1 = 0.$$

Let  $Q_+^{21}$  denote the projection onto  $\text{Ran } W_+(H_2, H_1)$ . Since  $1 - P_2 = (1 - P_2)(1 - Q_+^{21})$ , it suffices to prove

$$s\text{-}\lim_{t \rightarrow \infty} (1 - Q_+^{21}) e^{-itH_1} P_1 = 0.$$

We compute

$$\begin{aligned} \|(1 - Q_+^{21}) e^{-itH_1} P_1 \varphi\| &= \|e^{itH_2} (1 - Q_+^{21}) e^{-itH_1} P_1 \varphi\| \\ &= \|(1 - Q_+^{21}) e^{itH_2} e^{-itH_1} P_1 \varphi\| \\ &\rightarrow \|(1 - Q_+^{21}) W_+(H_2, H_1) \varphi\| = 0. \end{aligned}$$

The following result gives a criterion for completeness.

**Theorem 3.7.** *Assume that  $W_+(H_2, H_1)$  and  $W_+(H_1, H_2)$  both exist. Then*

$$W_+(H_1, H_2) = W_+(H_2, H_1)^*. \quad (3.15)$$

Furthermore,  $W_+(H_1, H_2)$  and  $W_+(H_2, H_1)$  are both complete.

**Remark 3.8.** The result (3.15) is nontrivial, since taking the adjoint is not a continuous operation in the strong operator topology.

*Proof.* Write  $W_{ij} = W_+(H_i, H_j)$  for simplicity. We get from (3.14)

$$W_{12} W_{21} = P_1, \quad \text{and} \quad W_{21} W_{12} = P_2.$$

Using (3.6)–(3.9), we find

$$W_{12} = P_1 W_{12} = W_{21}^* W_{21} W_{12} = W_{21}^* P_2 = W_{21}^*$$

and (3.15) is proved. The completeness of  $W_{21}$  now follows from

$$W_{21} W_{21}^* = W_{21} W_{12} = P_2,$$

and an analogous argument works for  $W_{12}$ .

The theory of wave operators can be localized to part of the spectrum of  $H_1$ . We state the definition:

#### DEFINITION 3.9

Let  $I_0$  be a Borel subset of  $\mathbf{R}$ . Then we define the local wave operators on  $I_0$  by

$$W_{\pm}(H_2, H_1; I_0) = s\text{-}\lim_{t \rightarrow \pm \infty} e^{itH_2} e^{-itH_1} E_{H_1}(I_0) P_1. \quad (3.16)$$

The definition of completeness is then changed to the condition  $\text{Ran}(W_{\pm}(H_2, H_1; I_0)) = E_{H_2}(I_0) \mathcal{H}_{\text{ac}}(H_2)$ . The results above are modified in an obvious manner.

### 3.2 The Cook–Kuroda criterion for existence of wave operators

The simplest criterion for existence of wave operators  $W_+(H_2, H_1)$  is discussed below. The first versions were given by Cook [7] and Kuroda [29]. Many refinements exist, but they will not be discussed here. The idea is simple: We write a continuously differentiable function as the integral of its derivative.

We recall that a subset  $\mathcal{D} \subset \mathcal{H}$  is called *fundamental*, if finite linear combinations of the vectors in  $\mathcal{D}$  form a dense subspace of  $\mathcal{H}$ .

**Theorem 3.10.** *Assume there exists a fundamental subset  $\mathcal{D}$  of  $\mathcal{H}_{\text{ac}}(H_1)$  with the following properties:*

- (i) *For each  $\varphi \in \mathcal{D}$  there exists  $t_0 = t_0(\varphi)$  such that for  $t \geq t_0$  we have  $\exp(-itH_1)\varphi \in \mathcal{D}(H_1) \cap \mathcal{D}(H_2)$ .*
- (ii)  *$(H_2 - H_1)\exp(-itH_1)\varphi$  is norm-continuous in  $t \in (t_0, \infty)$ .*
- (iii)  *$\|(H_2 - H_1)\exp(-itH_1)\varphi\| \in L^1((t_0, \infty))$ .*

*Then  $W_+(H_2, H_1)$  exists.*

*Proof.* Since  $\|\exp(itH_2)\exp(-itH_1)\| = 1$  for all  $t \in \mathbf{R}$ , it suffices to prove existence of  $W_+(H_2, H_1)$  on  $\mathcal{D}$ . Recalling the proof of the product rule for derivatives in elementary calculus, one can easily verify that for  $\varphi \in \mathcal{D}$



and, by assumption, this derivative is continuous in  $t, t \geq t_0$ . Thus for  $t_1, t_2 \geq t_0$

$$W(t_2)\varphi - W(t_1)\varphi = \int_{t_1}^{t_2} e^{isH_2} i(H_2 - H_1) e^{-isH_1} \varphi ds$$

and then

$$\|W(t_2)\varphi - W(t_1)\varphi\| \leq \int_{t_1}^{t_2} \|(H_2 - H_1)e^{-isH_1}\varphi\| ds.$$

Since the integrand is in  $L^1$ , the net  $\{W(t)\varphi\}$  is a Cauchy-net, and existence of the limit  $\lim_{t \rightarrow \infty} W(t)\varphi$  in  $\mathcal{H}$  follows.

One should note that verification of the hypotheses in this theorem can be difficult, depending on the information available on the operator  $H_1$  and the "interaction"  $H_2 - H_1$ .

#### 4. An abstract short range scattering theory

There are several abstract versions of scattering theory. They include the trace class scattering theory [5, 27, 38, 45] and the abstract theory developed by Kato and Kuroda [28, 30]. We present a simplified theory, which is strong enough to cover many interesting applications. The results presented here are distinguished by the introduction of an abstract short range condition. The proofs are time-dependent. These results were first given a paper with Mourre and Perry [23], without proof.

Let us recall the framework, and let us fix the notation.<sup>1</sup> Let  $\mathcal{H}$  be a separable Hilbert space and  $H_0$  a selfadjoint operator on  $\mathcal{H}$ . The domain of  $H_0$  is denoted  $\mathcal{D}(H_0)$ . We consider a perturbation  $V$  of  $H_0$ , called the potential, and the operator  $H$ , which is a selfadjoint extension of  $H_0 + V$ , defined suitably. Here we will not be concerned with technicalities and assume that  $V$  is a symmetric  $H_0$ -bounded operator with relative bound less than one. Then  $H = H_0 + V$  is the operator sum, and  $H$  is selfadjoint on  $\mathcal{D}(H) = \mathcal{D}(H_0)$ . The spectral family for  $H$  is denoted  $E_H$ . The subspace of absolute continuity is denoted  $\mathcal{H}_{ac}(H)$  and the singular subspace  $\mathcal{H}_s(H)$ . The corresponding projections are denoted  $P_{ac}(H)$  and  $P_s(H)$ , respectively. Let  $I_0 \subseteq \mathbf{R}$  be an open interval. We consider the scattering theory for the pair  $(H, H_0)$ , localized to the interval  $I_0$ . The two basic questions are:

(i) Existence of the wave operators

$$W_{\pm}(H, H_0; I_0) = s\text{-}\lim_{t \rightarrow \pm \infty} e^{itH} e^{-itH_0} E_{H_0}(I_0) P_{ac}(H_0). \quad (4.1)$$

(ii) Strong asymptotic completeness of the wave operators, i.e. the two results

$$\text{Ran}(W_{\pm}(H, H_0; I_0)) = E_H(I_0) P_{ac}(H) \mathcal{H}, \quad (4.2)$$

and

$$\sigma_{sc}(H) \cap I_0 = \emptyset. \quad (4.3)$$

In this theory we are able to prove a result stronger than (4.3), i.e. we prove

$$\sigma_s(H) \cap I_0 \text{ is discrete in } I_0. \quad (4.4)$$

This statement means that there is no singular continuous spectrum of  $H$  in  $I_0$  and, furthermore, the point spectrum of  $H$  in  $I_0$  consists at most of discrete eigenvalues with finite multiplicity. The only possible accumulation points of  $\sigma_{pp}(H) \cap I_0$  are the end points of  $I_0$ .

The abstract theory is based on the following two definitions:

#### DEFINITION 4.1

Let  $A$  be a selfadjoint operator on  $\mathcal{H}$ . We say that  $H_0$  satisfies propagation estimates with respect to  $A$  on  $I_0$ , if there exist real numbers  $s > s' > 1$  such that for all  $g \in C_0^\infty(I_0)$  the following two estimates hold:

$$\|(1 + A^2)^{-s/2} e^{-itH_0} g(H_0) (1 + A^2)^{-s/2}\| \leq c(1 + |t|)^{-s'} \quad \text{for all } t \in \mathbf{R}, \quad (4.5)$$

$$\|(1 + A^2)^{-s/2} e^{-itH_0} g(H_0) P_A^\pm\| \leq c(1 + |t|)^{-s'} \quad \text{for all } \pm t > 0. \quad (4.6)$$

Here  $P_A^+ = E_A((0, \infty))$  and  $P_A^- = 1 - P_A^+$ .

#### DEFINITION 4.2

Let  $A$  be a selfadjoint operator. The potential  $V$  is said to be a short range perturbation of  $H_0$  with respect to  $A$ , if

$$(H + i)^{-1} - (H_0 + i)^{-1} \text{ is a compact operator on } \mathcal{H}, \quad (4.7)$$

and if there exist a real number  $\mu > 1$  and integers  $j, k \geq 0$  such that the operator

$$(H + i)^{-j} V (H_0 + i)^{-k} (1 + A^2)^{\mu/2} \quad (4.8)$$

extends to a bounded operator on  $\mathcal{H}$ .

The abstract theorem given here is a version of the Enss method [11, 32, 36, 41]. The operator  $A$  is called an operator conjugate to  $H_0$ . See also Theorem 4.5.

**Theorem 4.3.** *Let  $H_0$ ,  $V$ ,  $I_0$  and  $H$  be as above. Assume that there exists a selfadjoint operator  $A$  such that  $H_0$  satisfies the propagation estimates with respect to  $A$  and such that the potential  $V$  is a short range perturbation of  $H_0$  with respect to  $A$ . Then the wave operators  $W_\pm(H, H_0; I_0)$  exist and are strongly asymptotically complete. Furthermore,  $\sigma_s(H) \cap I_0$  is discrete in  $I_0$ .*

*Proof.* The assumption (4.7) implies that  $g(H) - g(H_0)$  is compact for all  $g \in C_0^\infty(I_0)$ . Let us note that (4.5) implies  $I_0 \cap \sigma_s(H_0) = \emptyset$ , hence the projection  $P_{ac}(H_0)$  can be omitted in the computations. We consider only the  $+$  case, and start by proving existence of

$$\tilde{W} = s\text{-}\lim_{t \rightarrow \infty} (H + i)^{-j} e^{itH} e^{-itH_0} E_{H_0}(I_0). \quad (4.9)$$

By the Cook-Kuroda argument (see the proof of Theorem 3.10) it suffices to prove

$$\int_0^\infty \|(H+i)^{-j} V e^{-itH_0} g(H_0) \psi\| dt < \infty \quad (4.10)$$

for a dense set of  $\psi \in \mathcal{H}$ , and all  $g \in C_0^\infty(I_0)$ . Without loss of generality we can assume that  $s = \mu$ , with  $s$  from Definition 4.1 and  $\mu$  from Definition 4.2. Fix  $g \in C_0^\infty(I_0)$  and take  $\psi = (1 + A^2)^{-s/2} \varphi$ ,  $\varphi \in \mathcal{H}$  arbitrary. Using (4.5) and (4.8) we get

$$\begin{aligned} \|(H+i)^{-j} V e^{-itH_0} g(H_0) \psi\| &\leq \|(H+i)^{-j} V (H_0+i)^{-k} (1+A^2)^{s/2}\| \\ &\quad \cdot \|(1+A^2)^{-s/2} e^{-itH_0} (H_0+i)^k g(H_0) (1+A^2)^{-s/2}\| \cdot \|\varphi\| \\ &\leq c(1+|t|)^{-s'}, \end{aligned}$$

which is integrable, since  $s' > 1$ . To remove the factor  $(H+i)^{-j}$  in (4.9), we note (with an obvious notation)

$$\begin{aligned} e^{itH} e^{-itH_0} g(H_0) &= e^{itH} e^{-itH_0} (H_0+i)^{-j} \tilde{g}(H_0) \\ &= e^{itH} ((H_0+i)^{-j} - (H+i)^{-j}) e^{-itH_0} \tilde{g}(H_0) \\ &\quad + (H+i)^{-j} e^{itH} e^{-itH_0} \tilde{g}(H_0). \end{aligned}$$

Since  $(H_0+i)^{-j} - (H+i)^{-j}$  is compact by (4.7) and  $e^{-itH_0} \tilde{g}(H_0)$  tends to zero weakly as  $t \rightarrow \infty$ , the existence of  $W_+(H, H_0; I_0)$  follows.

We write  $W_+ = W_+(H, H_0; I_0)$  to simplify the notation. Let  $g \in C_0^\infty(I_0)$ . We prove that the operators

$$(W_\pm - 1)g(H_0)P_A^\pm \quad (4.11)$$

and

$$g(H)(W_\pm - 1)P_A^\pm \quad (4.12)$$

are compact. This is done by first proving that the operators

$$g_1(H)(W_\pm - 1)g_2(H_0)P_A^\pm \quad (4.13)$$

are compact for all  $g_1, g_2 \in C_0^\infty(I_0)$ . We have

$$\begin{aligned} g_1(H)(e^{itH} e^{-itH_0} - 1)g_2(H_0)P_A^\pm \\ = \int_0^t g_1(H) e^{i\tau H} i V e^{-i\tau H_0} g_2(H_0) P_A^\pm d\tau, \end{aligned} \quad (4.14)$$

where the integrand is norm continuous in  $\tau$  and compact, due to (4.7). We estimate for  $\tau > 0$

$$\begin{aligned} \|g_1(H) e^{i\tau H} i V e^{-i\tau H_0} g_2(H_0) P_A^\pm\| \\ \leq \|g_1(H) V (H_0+i)^{-k} (1+A^2)^{s/2}\| \cdot \|(1+A^2)^{-s/2} e^{-i\tau H_0} \tilde{g}_2(H_0) P_A^\pm\| \\ \leq c(1+|\tau|)^{-s'}. \end{aligned}$$

Thus the limit exists in norm as  $t \rightarrow \infty$ . An analogous result holds for  $t \rightarrow -\infty$ . Hence

intertwining relation for wave operators and the compactness of  $g(H) - g(H_0)$ . Thus compactness of the operators in (4.11) and (4.12) has been proved.

As the first step towards completeness we show that  $\sigma_s(H) \cap I_0$  is discrete in  $I_0$ . Let  $J \subset I_0$  be a relatively compact open interval and take  $g \in C_0^\infty(I_0)$ ,  $g(\lambda) = 1$  for all  $\lambda \in J$ . We recall from Theorem 3.2 the result

$$\text{Ran}(W_\pm) \subseteq \mathcal{H}_{\text{ac}}(H) = \mathcal{H}_s(H)^\perp,$$

hence  $P_s(H) W_\pm = 0$ .

$$\begin{aligned} P_s(H) E_H(J) &= P_s(H) E_H(J) g(H) \\ &= P_s(H) E_H(J) g(H) (P_A^+ + P_A^-) \\ &= P_s(H) E_H(J) g(H) (1 - W_+) P_A^+ \\ &\quad + P_s(H) E_H(J) g(H) (1 - W_-) P_A^-. \end{aligned}$$

By the compactness of (4.12) we conclude that this operator is compact, hence of finite rank, which proves the last statement in the theorem.

The final step in the proof of completeness consists in proving

$$\text{Ran}(W_\pm) = E_H(I_0) \mathcal{H}_{\text{ac}}(H).$$

Since  $I_0 \setminus \sigma_{\text{pp}}(H)$  is open by the first part of the proof, we can take  $g \in C_0^\infty(I_0 \setminus \sigma_{\text{pp}}(H))$ . By general results on wave operators given in § 3 (see in particular Theorem 3.7) it suffices to prove

$$s\text{-}\lim_{t \rightarrow \pm \infty} e^{itH_0} e^{-itH} g(H) = W_\pm^* g(H). \quad (4.15)$$

Take  $\psi \in \mathcal{H}_{\text{ac}}(H)$  with  $\psi = g(H)\psi$  and compute as follows in the  $+$ -case:

$$\begin{aligned} \|e^{itH_0} e^{-itH} g(H)\psi - W_+^* g(H)\psi\| &= \|e^{-itH} g(H)\psi - e^{-itH_0} W_+^* g(H)\psi\| \\ &= \|(P_A^+ + P_A^-)(e^{-itH} g(H)\psi - e^{-itH_0} W_+^* g(H)\psi)\| \\ &\leq A_+(t) + A_-(t) \end{aligned}$$

with an obvious notation.

$$\begin{aligned} A_+(t) &= \|P_A^+(e^{-itH} g(H)\psi - e^{-itH_0} W_+^* g(H)\psi)\| \\ &= \|P_A^+(1 - W_+^*)g(H)e^{-itH}\psi\|, \end{aligned}$$

where we used the intertwining relation. By (4.12) the operator  $P_A^+(1 - W_+^*)g(H)$  is compact, and  $e^{-itH}\psi \rightarrow 0$  as  $t \rightarrow \infty$  weakly, so  $A_+(t) \rightarrow 0$  as  $t \rightarrow \infty$ .

$$A_-(t) \leq \|P_A^- e^{-itH} g(H)\psi\| + \|P_A^- e^{-itH_0} W_+^* g(H)\psi\|. \quad (4.16)$$

The estimate in (4.6) implies

$$s\text{-}\lim_{t \rightarrow \infty} P_A^- e^{-itH_0} g(H_0) = 0.$$

Since  $W_+^* g(H)\psi = g(H_0) W_+^* \psi$ , the second term in (4.16) tends to zero.

$$\begin{aligned} \|P_A^- e^{-iH} g(H)\psi\| &\leq \|P_A^- W_-^* e^{-iH} g(H)\psi\| \\ &\quad + \|P_A^- (1 - W_-^*) e^{-iH} g(H)\psi\|. \end{aligned}$$

These two terms tend to zero, as above. This concludes the proof of Theorem 4.3.

*Remark 4.4.* (i) Note that we need the existence of both  $W_+$  and  $W_-$  to prove the completeness of  $W_+$ .

(ii) One can state the result in a form that does not introduce the potential. Instead, one assumes the existence of two selfadjoint operators  $H_0$  and  $H$  satisfying (4.7). The assumption (4.8) is reformulated as: For some integer  $k \geq 0$  and a real number  $\mu > 1$  the operator

$$E_H(I_0)(H - H_0)(H_0 + i)^{-k}(1 + A^2)^{\mu/2}$$

extends to a bounded operator on  $\mathcal{H}$ . One could also use the stronger condition

$$((H + i)^{-k-1} - (H_0 + i)^{-k-1})(1 + A^2)^{\mu/2}$$

extends to a bounded operator on  $\mathcal{H}$ .

To complete the abstract theory one can give other conditions which are simpler to verify than the propagation estimates in Definition 4.1. There is a method based on commutators due to Eric Mourre [33, 23]. We state the result (see [18, 23]) in the following form:

**Theorem 4.5.** *Let  $H_0$  and  $A$  be selfadjoint operators on a Hilbert space  $\mathcal{H}$ . Let  $\lambda_0 \in \mathbf{R}$ . We suppose:*

- (a)  $\mathcal{D}(A) \cap \mathcal{D}(H_0)$  is a core for  $H_0$ .
- (b)  $e^{i\theta A} \mathcal{D}(H_0) \subseteq \mathcal{D}(H_0)$  and for each  $\psi \in \mathcal{D}(H_0)$  we have

$$\sup_{|\theta| \leq 1} \|H_0 e^{i\theta A} \psi\| < \infty.$$

- (c) *The commutator  $i[H_0, A]$ , defined as a form on  $\mathcal{D}(A) \cap \mathcal{D}(H_0)$ , is bounded below and closable. The selfadjoint operator associated with its closure is denoted  $iB_1$ . Assume  $\mathcal{D}(B_1) \supseteq \mathcal{D}(H_0)$ . Assume inductively for  $j = 2, 3, \dots$  that the form  $i[iB_{j-1}, A]$  is bounded below and closable. The associated operator is denoted  $iB_j$ . Assume  $\mathcal{D}(B_j) \supseteq \mathcal{D}(H_0)$ .*
- (d) *There exist  $\alpha > 0$ ,  $\delta > 0$ , and a compact operator  $K$  such that with  $J = (\lambda_0 - \delta, \lambda_0 + \delta)$  the Mourre estimate*

$$E_{H_0}(J) iB_1 E_{H_0}(J) \geq \alpha E_{H_0}(J) + K \quad (4.17)$$

*holds.*

*Then we have*

- (i)  $\sigma_s(H_0) \cap J$  is discrete in  $J$ .
- (ii) *The estimates in (4.5) and (4.6) hold with  $I_0 = J \setminus \sigma_s(H_0)$  for all  $s > s' > 0$ .*

We refer to the original papers for the proofs [23, 33]. The proofs are quite long.

In the next section we will verify the conditions in Definitions 4.1 and 4.2 by direct computation.

## 5. Application to $H_0 = -\Delta$

In this section we verify the conditions in the abstract theory in §4 by direct commutator computations in the case  $H_0 = -\Delta$ . The results are an expanded version of some results in [22].

Let  $H_0 = -\Delta$  on  $\mathcal{H} = l^2(\mathbf{R}^d)$ . It is a selfadjoint operator with domain  $\mathcal{D}(H_0) = H^2(\mathbf{R}^d)$ , the usual Sobolev space of order 2. See [37, Section IX.7] for the results needed here. Let  $A = \frac{1}{2i}(x \cdot \nabla + \nabla \cdot x)$  denote the generator of dilations. We state the result in the following form:

**Theorem 5.1.** *The operators  $H_0$  and  $A$  above satisfy the estimates (4.5) and (4.6) with  $I_0 = (0, \infty)$  for all  $s > s' > 0$*

*Proof.* Define  $L_0 = \log H_0$  using the functional calculus. A straightforward calculation proves

$$i[L_0, A] = 2 \cdot 1.$$

Therefore we have

$$e^{itL_0} A e^{-itL_0} = A + 2t \cdot 1,$$

which implies three propagation estimates for  $L_0$ :

$$\|(1 + A^2)^{-s/2} e^{-itL_0} (1 + A^2)^{-s'/2}\| \leq c(1 + |t|)^{-s} \quad \text{for all } t \in \mathbf{R}. \quad (5.1)$$

$$\|(1 + A^2)^{-s/2} e^{-itL_0} P_A^\pm\| \leq c(1 + |t|)^{-s} \quad \text{for all } \pm t > 0. \quad (5.2)$$

$$P_A^\mp e^{-itL_0} P_A^\pm = 0 \quad \text{for all } \pm t > 0. \quad (5.3)$$

By the functional calculus

$$e^{-itL_0} = H_0^{-it}. \quad (5.4)$$

The propagation estimates for  $H_0$  are now derived using these results and the one-dimensional Mellin transform in momentum space. We recall the definitions:  $\mathcal{M}: L^2(\mathbf{R}_+) \rightarrow L^2(\mathbf{R})$  is given by

$$(\mathcal{M}f)(\lambda) = \int_0^\infty f(p) p^{-i\lambda-1} dp$$

with inverse

$$(\mathcal{M}^{-1}g)(p) = \frac{1}{2\pi} \int_{-\infty}^\infty g(\lambda) p^{i\lambda} d\lambda.$$

Let  $g \in C_0^\infty(\mathbf{R}_+)$ . In momentum space  $e^{-itH_0}g(H_0)$  is multiplication by the function  $e^{-it\rho}g(\rho)$ ,  $\rho = \rho^2$ . Using the Mellin transform (shifting a  $2\pi$  factor) we write

$$e^{-it\rho}g(\rho) = \int_{-\infty}^{\infty} G_t(\lambda)\rho^{i\lambda}d\lambda$$

with

$$G_t(\lambda) = \frac{1}{2\pi} \int_0^\infty e^{-it\rho}g(\rho)\rho^{-i\lambda-1}d\rho.$$

*Lemma 5.2. The function  $G_t(\lambda)$  satisfies the following estimates:*

$$|G_t(\lambda)| \leq C_N |t|^{-N} (1 + |\lambda|)^N \quad \text{for all } N \geq 1, \quad (5.5)$$

and for  $t \cdot \lambda > 0$

$$|G_t(\lambda)| \leq C_N (1 + |t + \lambda|)^{-N} \quad \text{for all } N \geq 1. \quad (5.6)$$

*Proof.* The results follow from straightforward (non-) stationary phase computations. Let us give some of the details for (5.6). Assume  $\lambda > 0$ ,  $t > 0$ , and  $\text{supp } g \subset (c_1, c_2)$ ,  $0 < c_1 < c_2 < \infty$ . We have for any  $N \geq 1$

$$e^{-it\rho}\rho^{-i\lambda} = \left( \frac{i}{(\lambda/\rho) + t} \cdot \frac{d}{dt} \right)^N e^{-it\rho - i\lambda \log \rho}.$$

Integration by parts yields

$$G_t(\lambda) = c \int_0^\infty e^{-it\rho - i\lambda \log \rho} \left( \frac{d}{dt} \cdot \frac{i}{(\lambda/\rho) + t} \right)^N (g(\rho)\rho^{-1}) d\rho.$$

For  $\rho \in (c_1, c_2)$  we have

$$\frac{1}{(\lambda/c_1) + t} \leq \frac{1}{(\lambda/\rho) + t} \leq \frac{1}{(\lambda/c_2) + t}$$

Now we have

$$\frac{d}{d\rho} \frac{1}{(\lambda/\rho) + t} = \frac{\lambda}{\rho^2} \left( \frac{\lambda}{\rho} + t \right)^{-2} \leq \frac{1}{c_1} \frac{1}{(\lambda/c_2) + t}$$

and in general, as one easily verifies,

$$\left| \left( \frac{d}{d\rho} \right)^j \frac{1}{(\lambda/\rho) + t} \right| \leq C_j ((\lambda + t)^{-j} + (\lambda + t)^{-1}),$$

for  $j = 1, 2, \dots, N$ . Thus we get an estimate

$$|G_t(\lambda)| \leq C |1 + \lambda + t|^{-N}$$

which concludes the proof of the lemma.

We can now prove the propagation estimates for  $H_0$ . By the functional calculus

we have

$$e^{-itH_0}g(H_0) = \int_{-\infty}^{\infty} G_t(\lambda)H_0^{i\lambda}d\lambda.$$

Using (5.1), (5.4), and (5.5) we obtain for  $s > 1$

$$\begin{aligned} \|(1+A^2)^{-s/2}e^{-itH_0}g(H_0)(1+A^2)^{-s/2}\| &\leq C \int_{-\infty}^{\infty} |G_t(\lambda)|(1+|\lambda|)^{-s}d\lambda \\ &\leq C_{N,s}|t|^{-N}, \end{aligned}$$

if  $N$  is an integer satisfying  $N < s - 1$ . Thus the estimate holds for all  $s$  and  $N$  satisfying this condition, and the general result (4.5) follows by complex interpolation.

We prove (4.6) in the  $t > 0$  case. Write

$$(1+A^2)^{-s/2}e^{-itH_0}g(H_0)P_A^+ = \int_{-\infty}^{\infty} G_t(\lambda)(1+A^2)^{-s/2}H_0^{i\lambda}P_A^+d\lambda.$$

Then the integral over negative  $\lambda$  is estimated using (5.2) and (5.5) for  $N < s - 1$

$$\begin{aligned} \left\| \int_{-\infty}^0 G_t(\lambda)(1+A^2)^{-s/2}H_0^{i\lambda}P_A^+d\lambda \right\| &\leq c \int_{-\infty}^0 |G_t(\lambda)|(1+|\lambda|)^{-s}d\lambda \\ &\leq C_{N,s}|t|^{-N}. \end{aligned}$$

For the integral over positive  $\lambda$  we use the estimate (5.6) to get

$$\begin{aligned} \left\| \int_0^{\infty} G_t(\lambda)(1+A^2)^{-s/2}H_0^{i\lambda}P_A^+d\lambda \right\| &\leq c \int_0^{\infty} |G_t(\lambda)|d\lambda \\ &\leq c \int_0^{\infty} (1+t+\lambda)^{-m}d\lambda \\ &\leq ct^{-m+1} \end{aligned}$$

for any  $m > 1$ . The proof is now completed using complex interpolation.

*Remark 5.3.* The Mellin transform was used by Perry [35] in a different manner to derive the basic propagation estimate used in the Enss method.

We must now characterize more explicitly which potentials  $V(x)$  are short range in this abstract framework. The results obtained here cover the well-known short range conditions. In rough asymptotic form the short range condition is  $V(x) = O(|x|^{-1-\varepsilon})$ ,  $\varepsilon > 0$ , as  $|x| \rightarrow \infty$  for a multiplicative potential. A precise statement for a general (not necessarily multiplicative) potential is given in the following proposition:

#### PROPOSITION 5.4.

*Let  $V$  be a symmetric  $H_0$ -compact operator such that for some  $\mu > 1$  the operator  $V(H_0 + i)^{-1}(1+x^2)^{\mu/2}$  extends to a bounded operator on  $L^2(\mathbf{R}^d)$ . Then  $V$  is a short range perturbation of  $H_0 = -\Delta$  with respect to  $A = 1/(2i)(\nabla \cdot x + x \cdot \nabla)$ .*



*Proof.* The assumptions imply (4.7) immediately, and a simple complex interpolation argument shows that  $V(H_0 + i)^{-2}(1 + A^2)^{\mu/2}$  extends to a bounded operator on  $L^2(\mathbf{R}^d)$ . Thus the conditions in Definition 4.2 are verified.

More interesting is to note that some classes of oscillatory potentials also can be shown to fit into the framework given here. There is already a considerable literature [6, 36, 42] on this subject, so we only give one representative result. The result is not new, see for example [41, Note added in proof], and for further references and results, [42]. Note also that our technique allows us to treat general non-central oscillatory potentials.

### PROPOSITION 5.5.

Let  $\alpha > 0$ ,  $\beta > 0$ ,  $2\alpha + \beta > 3$ ,  $b, c \in \mathbf{R}$ , and define for  $x \in \mathbf{R}^d$ ,  $x \neq 0$ ,

$$V(x) = c \frac{\sin(b|x|^\alpha)}{|x|^\beta}.$$

Then  $V$  is a short range perturbation of  $H_0 = -\Delta$  with respect to  $A = 1/(2i)(\nabla \cdot x + x \cdot \nabla)$ .

*Proof.* The proof is based on the following lemma:

*Lemma 5.6.* Let  $g: \mathbf{R}^d \rightarrow \mathbf{R}^d$  be a real-valued continuously differentiable function. Let  $U$  and  $W$  be bounded real-valued functions such that for some integer  $k$  and real number  $\mu > 1$  the operators

$$g_j(x)U(x)(H_0 + 1)^{-k}(1 + A^2)^{\mu/2}, \quad j = 1, 2, \dots, d,$$

$$(\nabla \cdot g(x))U(x)(H_0 + 1)^{-k}(1 + A^2)^{\mu/2},$$

and

$$W(H_0 + 1)^{-k}(1 + A^2)^{\mu/2}$$

extend to bounded operators on  $L^2(\mathbf{R}^d)$ . Let

$$V(x) = [g(x) \cdot \nabla, U(x)] + W(x).$$

Then the operator

$$(H_0 + 1)^{-1/2} V (H_0 + 1)^{-k-1/2} (1 + A^2)^{\mu/2}$$

extends to a bounded operator on  $L^2(\mathbf{R}^d)$ .

*Proof.* We compute:

$$\begin{aligned} & (H_0 + 1)^{-1/2} V (H_0 + 1)^{-k-1/2} (1 + A^2)^{\mu/2} \\ &= ((H_0 + 1)^{-1/2} \nabla) \cdot (gU) (H_0 + 1)^{-k-1/2} (1 + A^2)^{\mu/2} \\ &+ (H_0 + 1)^{-1/2} (\nabla \cdot g) U (H_0 + 1)^{-k-1/2} (1 + A^2)^{\mu/2} \\ &- (H_0 + 1)^{-1/2} (gU) (H_0 + 1)^{-k} \cdot (\nabla (H_0 + 1)^{-1/2}) (1 + A^2)^{\mu/2} \\ &+ (H_0 + 1)^{-1/2} W (H_0 + 1)^{-k-1/2} (1 + A^2)^{\mu/2}. \end{aligned}$$

Since

$$(1 + A^2)^{-\mu/2} (H_0 + 1)^{-1/2} (1 + A^2)^{\mu/2}$$

and

$$(1 + A^2)^{-\mu/2} \nabla (H_0 + 1)^{-1/2} (1 + A^2)^{\mu/2}$$

extend to bounded operators, the result follows.

Using this lemma we can prove the proposition. It suffices to consider the case  $b = c = 1$ . Assume  $\alpha > 0$  and  $\beta > 0$ . Compute the commutator

$$\left[ \frac{1}{\alpha r^{\alpha-1}} \frac{\partial}{\partial r}, \frac{\sin(r^\alpha)}{r^\beta} \right] = \frac{\cos(r^\alpha)}{r^\beta} - \frac{\beta \sin(r^\alpha)}{\alpha r^{\alpha+\beta}}. \quad (5.7)$$

Assume  $\alpha + \beta > 2$ . Using the lemma, we conclude that

$$\frac{\cos(r^\alpha)}{r^\beta}$$

is short range, and similarly for

$$\frac{\sin(r^\alpha)}{r^\beta}.$$

Now use these results in (5.7) and repeat the computations in the proof of the lemma with an extra factor  $(H_0 + 1)^{-1/2}$  on the left to conclude that

$$\frac{\cos(r^\alpha)}{r^\beta} \quad \text{and} \quad \frac{\sin(r^\alpha)}{r^\beta}$$

are short range for  $2\alpha + \beta > 3$ . This concludes the proof.

*Remark 5.7.* In [22] it is claimed that one can iterate the arguments above indefinitely. This leads to the result that

$$(H_0 + i)^{-j} V(H_0 + i)^{-k} (1 + A^2)^{\mu/2}$$

is bounded on  $L^2(\mathbf{R}^d)$ , provided  $(\alpha - 1)j + \beta > 1$  and  $\mu$  and  $k$  are chosen appropriately. However, we cannot replace  $(H_0 + i)^{-j}$  by  $(H + i)^{-j}$  for  $j > 1$ , since  $(H + i)^j (H_0 + i)^{-j}$  is not bounded. Thus the result [22, Proposition 5.2] is not correct.

## 6. Stark Hamiltonians

In this section we discuss the Stark Hamiltonians in detail. We derive the basic results in the general  $d$ -dimensional case in the first subsection. In the second subsection the one-dimensional case is further investigated, and then in the last subsection we discuss classical scattering in the one-dimensional case. This leads to a discrepancy between classical and quantum scattering, which is resolved in § 7.

The results in this section have been taken from [19, 20, 21, 26, 25].

## 6.1 General results on quantum scattering

The operator

$$H_0 = -\Delta + F \cdot x, \quad F \in \mathbf{R}^d, \quad F \neq 0,$$

is called the free *Stark* Hamiltonian. This operator is essentially selfadjoint on both  $C_0^\infty(\mathbf{R}^d)$  and  $\mathcal{S}(\mathbf{R}^d)$ , the Schwartz space. It is easy to show that  $\sigma(H_0) = \sigma_{ac}(H_0) = \mathbf{R}$ . It represents the energy of an electron moving in a constant electric field given by  $F$ . The model is somewhat unphysical in the sense that the spectrum is unbounded below. Mathematically it is a very interesting operator, and, as we shall see, many of the difficulties come precisely from the fact that the operator  $H_0$  is unbounded below.

We will normalize the choice of  $F$  to  $F = (1, 0, \dots, 0)$ , hence here

$$H_0 = -\Delta + x_1. \quad (6.1)$$

We identify  $\mathbf{R}^d = \mathbf{R} \times \mathbf{R}^{d-1}$  and write  $x = (x_1, x_\perp)$ ,  $x_1 \in \mathbf{R}$ ,  $x_\perp \in \mathbf{R}^{d-1}$ . We use the notation  $p = -i\nabla$  and  $p_j = -i\partial/\partial x_j$ ,  $j = 1, \dots, d$ . Thus  $H_0 = p^2 + x_1$ .

The one-dimensional Stark Hamiltonian is diagonalized as follows:

$$e^{-ip_1^3/3}(p_1^2 + x_1)e^{ip_1^3/3} = x_1. \quad (6.2)$$

Thus we have in the general  $d$ -dimensional case

$$e^{-ip_1^3/3}H_0e^{ip_1^3/3} = x_1 - \Delta_{x_\perp}.$$

Using this representation one verifies the results on essential selfadjointness and type of spectrum mentioned above.

To apply the abstract theory in §4 we need a conjugate operator  $A$ . Here we choose  $A = -p_1$ . We then get  $i[H_0, A] = 1$  on  $\mathcal{S}(\mathbf{R}^d)$ . The computations in the proof of Theorem 5.1 are valid here, since

$$e^{itH_0}Ae^{-itH_0} = A + t \cdot 1,$$

so we have for any  $s \geq 0$

$$\|(1 + A^2)^{-s/2}e^{-itH_0}(1 + A^2)^{-s/2}\| \leq c(1 + |t|)^{-s} \quad \text{for all } t \in \mathbf{R}, \quad (6.3)$$

$$\|(1 + A^2)^{-s/2}e^{-itH_0}P_A^\pm\| \leq c(1 + |t|)^{-s} \quad \text{for all } \pm t > 0, \quad (6.4)$$

$$P_A^\mp e^{-itH_0}P_A^\pm = 0 \quad \text{for all } \pm t > 0. \quad (6.5)$$

Now we need to insert  $g(H_0)$ ,  $g \in C_0^\infty(\mathbf{R})$ , in the estimates (6.3) and (6.4). We claim that

$$(1 + A^2)^{s/2}g(H_0)(1 + A^2)^{-s/2} \quad (6.6)$$

is bounded for all  $s \geq 0$ . Let  $\hat{g}$  denote the Fourier transform of  $g \in C_0^\infty(\mathbf{R})$ . Then

$$g(H_0) = (2\pi)^{-1/2} \int e^{itH_0} \hat{g}(\tau) d\tau.$$

Write  $\langle A \rangle = (1 + A^2)^{1/2}$ . Now we estimate

$$\|\langle A \rangle^s e^{itH_0} \langle A \rangle^{-s}\| = \|\langle A \rangle^s e^{itH_0} \langle A \rangle^{-s} e^{-itH_0}\| = \|\langle A \rangle^s \langle A + \tau \rangle^{-s}\| \leq c \langle \tau \rangle^s$$

and  $\langle \tau \rangle^s \hat{g} \in L^1(\mathbf{R})$  for any  $s \geq 0$ , so the result (6.6) follows.

We summarize these computations in a proposition:

### PROPOSITION 6.1

The propagation estimates (4.5) and (4.6) hold for  $H_0 = -\Delta + x_1$  and  $A = i\partial/\partial x_1$  on  $L^2(\mathbf{R}^d)$ , with  $s' = s \geq 0$ .

It remains to characterize more explicitly which potentials are short range for  $H_0$  with respect to  $A$ . Some preparation is needed. We start with two lemmas. First we fix a function  $\chi_+ \in C^\infty(\mathbf{R})$  with the properties  $0 \leq \chi_+ \leq 1$ ,  $\chi_+(x_1) = 1$  for  $x_1 \geq 2$ ,  $\chi_+(x_1) = 0$  for  $x_1 \leq 1$ , and we let  $\chi_- = 1 - \chi_+$ . We have the following result from [15, 41].

*Lemma 6.2.* Let  $z \in \mathbb{C}$ ,  $\text{Im } z \neq 0$ . Then the following operators are bounded on  $L^2(\mathbf{R}^d)$ .

$$\chi_+(x_1) \Delta (H_0 - z)^{-1}, \quad (6.7)$$

$$\chi_+(x_1) x_1 (H_0 - z)^{-1}, \quad (6.8)$$

$$\chi_+(x_1) \sqrt{x_1} p_j (H_0 - z)^{-1}, \quad j = 1, \dots, d. \quad (6.9)$$

*Proof.* We recall the proof from [15, 41]. Computing as quadratic forms on  $\mathcal{S}(\mathbf{R}^d)$ , we get (writing  $\chi_+$  instead of  $\chi_+(x_1)$ )

$$\begin{aligned} (\chi_+(p^2 + x_1))^* (\chi_+(p^2 + x_1)) &= (p^2 + x_1) \chi_+^2(p^2 + x_1) \\ &= p^2 \chi_+^2 p^2 + x_1^2 \chi_+^2 + x_1 \chi_+^2 p^2 + p^2 x_1 \chi_+^2 \\ &= p^2 \chi_+^2 p^2 + x_1^2 \chi_+^2 + 2 \sum_{j=1}^d p_j x_1 \chi_+^2 p_j + \\ &\quad [p_1, [p_1, x_1 \chi_+^2]]. \end{aligned}$$

We observe

$$[p_1, [p_1, x_1 \chi_+^2]] = -\frac{d^2}{dx_1^2} (x_1 \chi_+^2) \geq -c$$

for some  $c > 0$ . For  $\varphi \in \mathcal{S}(\mathbf{R}^d)$  we then get

$$\|\chi_+ H_0 \varphi\|^2 + c \|\varphi\|^2 \geq \|\chi_+ p^2 \varphi\|^2 + \|x_1 \chi_+ \varphi\|^2 + 2 \sum_{j=1}^d \|\sqrt{x_1} \chi_+ p_j \varphi\|^2.$$

Since  $H_0$  maps  $\mathcal{S}(\mathbf{R}^d)$  into  $\mathcal{S}(\mathbf{R}^d)$ , we can take  $\varphi = (H_0 - z)^{-1} \psi$ ,  $\psi \in \mathcal{S}(\mathbf{R}^d)$ , and the lemma follows.

We introduce the weight function

$$w(x_1) = \chi_+(x_1) + (1 + x_1^2)^{-1/2} \chi_-(x_1).$$

**Lemma 6.3.** Let  $0 \leq \delta \leq 1$ . Then the operator

$$(1 + A^2)^\delta (H_0 + i)^{-1} w(x_1)^\delta$$

extends to a bounded operator on  $L^2(\mathbf{R}^d)$ .

*Proof.* Using complex interpolation we see that it suffices to prove the result for  $\delta = 1$ . (The case  $\delta = 0$  is trivial.) We start by noting

$$\|p_1^2(H_0 + i)^{-1} w(x_1) \varphi\| \leq \|p^2(H_0 + i)^{-1} w(x_1) \varphi\|.$$

Then

$$p^2(H_0 + i)^{-1} w(x_1) = (H_0 - x_1)(H_0 + i)^{-1} w(x_1).$$

Since  $H_0(H_0 + i)^{-1}$  is bounded, it suffices to look at the  $x_1$ -term. We compute the commutator

$$[x_1, (H_0 + i)^{-1}] = -2i(H_0 + i)^{-2} p_1 + 2(H_0 + i)^{-3}.$$

Hence

$$\begin{aligned} x_1(H_0 + i)^{-1} w(x_1) &= (H_0 + i)^{-1} x_1 \chi_+(x_1) + (H_0 + i)^{-1} x_1 \chi_-(x_1) (1 + x_1^2)^{-1/2} \\ &\quad - 2i(H_0 + i)^{-2} p_1 \chi_+(x_1) - 2i(H_0 + i)^{-2} p_1 \chi_-(x_1) (1 + x_1^2)^{-1/2} \\ &\quad + 2(H_0 + i)^{-3} w(x_1). \end{aligned}$$

The first term is bounded by (6.8). The second and fifth terms are trivially bounded. The third term is bounded by (6.9). The fourth term requires some computations. Write  $\tilde{\chi}_- = \chi_-(x_1)(1 + x_1^2)^{-1/2}$ . Since the commutator  $[p_1, \tilde{\chi}_-]$  is bounded, it suffices to consider  $p_1 \tilde{\chi}_- (H_0 + i)^{-1}$ . We compute as in the proof of Lemma 6.2. Let  $\varphi \in \mathcal{S}(\mathbf{R}^d)$ .

$$\begin{aligned} \|p_1 \tilde{\chi}_- (H_0 + i)^{-1} \varphi\|^2 &= ((H_0 + i)^{-1} \varphi, \tilde{\chi}_- p_1^2 \tilde{\chi}_- (H_0 + i)^{-1} \varphi) \\ &\leq ((H_0 + i)^{-1} \varphi, \tilde{\chi}_- p^2 \tilde{\chi}_- (H_0 + i)^{-1} \varphi) \\ &= ((H_0 + i)^{-1} \varphi, \{\tilde{\chi}_-^2 p^2 - 2i\tilde{\chi}_- p_1 \tilde{\chi}_- + \tilde{\chi}_- \tilde{\chi}_-''\} \\ &\quad (H_0 + i)^{-1} \varphi) \\ &= ((H_0 + i)^{-1} \varphi, \tilde{\chi}_-^2 (H_0 - x_1)(H_0 + i)^{-1} \varphi) \\ &\quad + ((H_0 + i)^{-1} \varphi, \tilde{\chi}_- \tilde{\chi}_-'' (H_0 + i)^{-1} \varphi) \\ &\quad - 2i(p_1 \tilde{\chi}_- (H_0 + i)^{-1} \varphi, \tilde{\chi}_- (H_0 + i)^{-1} \varphi) \\ &\leq c \|\varphi\|^2 + \frac{1}{2} \|p_1 \tilde{\chi}_- (H_0 + i)^{-1} \varphi\|^2. \end{aligned}$$

We see that  $p_1 \tilde{\chi}_- (H_0 + i)^{-1}$  extends to a bounded operator on  $L^2(\mathbf{R}^d)$ . This concludes the proof of the lemma.

With this preparation the main result is easily obtained. We state the following theorem:

**Theorem 6.4.** Let  $V$  be a symmetric operator on  $L^2(\mathbf{R}^d)$  with the properties  $\mathcal{D}(V) \supseteq \mathcal{D}(H_0)$ ,  $V(H_0 + i)^{-1}$  is compact, and for some  $\mu > 1$  the operator

$$w(x_1)^{-\mu/2} V(H_0 + i)^{-1}$$

extends to a bounded operator on  $L^2(\mathbf{R}^d)$ . Then  $V$  is a short range perturbation of  $H_0$

with respect to  $A$ . Thus the wave operators  $W_{\pm}(H_0 + V, H_0)$  exist and are strongly asymptotically complete. Furthermore, the point spectrum of  $H_0 + V$  is discrete in  $\mathbf{R}$ .

*Proof.* We write  $H = H_0 + V$ .  $H$  is selfadjoint with domain  $\mathcal{D}(H) = \mathcal{D}(H_0)$ . By assumption the operator  $(H + i)^{-1} - (H_0 + i)^{-1}$  is compact. Furthermore,

$$(H + i)^{-1} V (H_0 + i)^{-1} (1 + A^2)^{\mu/2} = ((H + i)^{-1} (H_0 + i)) \cdot ((H_0 + i)^{-1} V w(x_1)^{-\mu/2}) \cdot (w(x_1)^{\mu/2} (H_0 + i)^{-1} (1 + A^2)^{\mu/2}).$$

The first factor is bounded, since  $\mathcal{D}(H) = \mathcal{D}(H_0)$ . The adjoint of the second factor is bounded by assumption, hence the second factor is bounded. The third factor is bounded by Lemma 6.3, since we can assume  $\mu \leq 2$  without loss of generality. Thus the conditions (4.7) and (4.8) in Definition 4.2 have been verified, and the results follow from Theorem 4.3.

Let us briefly discuss this result. First one should note that we are not requiring  $V$  to be a multiplication operator. It is easy to give examples of finite rank operators satisfying our assumptions. If one assumes that  $V$  is multiplication by a function  $V(x) = V(x_1, x_{\perp})$ , then the requirements imposed are in rough asymptotic terms the following:

$$V(x_1, x_{\perp}) = \begin{cases} O(|x_1|^{-1/2-\varepsilon}) & \text{as } x_1 \rightarrow -\infty, \\ o(|x_1|) & \text{as } x_1 \rightarrow \infty, \\ o(1) & \text{as } |x_{\perp}| \rightarrow \infty. \end{cases}$$

Previous results on the scattering theory for Stark Hamiltonians include the results by Avron-Herbst [4], Herbst [14], and Yajima [46]. The above result covers all these previous results.

## 6.2 One-dimensional quantum scattering

In the one-dimensional case there are other classes of potentials which are short range perturbations of  $H_0$  with respect to  $A$ . We give some results here. See [20, 21] for further results. We denote by  $\mathcal{B}^k(\mathbf{R})$  the realvalued functions on  $\mathbf{R}$  which are  $k$  times continuously differentiable, such that the function and all its derivatives up to order  $k$  are bounded on  $\mathbf{R}$ . If  $V$  is a bounded realvalued function on  $\mathbf{R}$ , then  $H = H_0 + V$  is selfadjoint on  $\mathcal{D}(H_0)$ .

*Lemma 6.5.* Suppose  $V$  is a realvalued function such that for some  $W \in \mathcal{B}^4(\mathbf{R})$  we have  $V = W''$ . Then the two operators

$$(H + i)^{-1} V p_1 (H_0 + i)^{-1} \tag{6.10}$$

and

$$(H + i)^{-2} V p_1^2 (H_0 + i)^{-2} \tag{6.11}$$

extend to bounded operators on  $L^2(\mathbf{R})$ .

*Proof.* Let  $U \in \mathcal{B}^2(\mathbf{R})$ . As a quadratic form on  $\mathcal{S}(\mathbf{R}) \times \mathcal{S}(\mathbf{R})$  we have

$$i[H_0, U] = 2U' p_1 - iU''.$$

Computing in the sense of quadratic form on  $\mathcal{S}(\mathbf{R}) \times \mathcal{S}(\mathbf{R})$ , omitting the vectors to simplify the notation, we find

$$\begin{aligned} \frac{d}{ds} e^{isH} U e^{-isH_0} &= i e^{isH} (HU - UH_0) e^{-isH_0} \\ &= e^{isH} (iVU + i[H_0, U]) e^{-isH_0} \\ &= e^{isH} (iVU + 2U'p_1 - iU'') e^{-isH_0}. \end{aligned} \quad (6.12)$$

Using this relation we derive

$$\begin{aligned} 2(H+i)^{-1} e^{isH} U' p_1 e^{-isH_0} (H_0+i)^{-1} \\ = (H+i)^{-1} e^{isH} (iHU - iUH_0 - iVU + iU'') e^{-isH_0} (H_0+i)^{-1}. \end{aligned}$$

Now we see that the right hand side is a bounded operator on  $L^2(\mathbf{R})$ , hence the left hand operator extends to a bounded operator on  $L^2(\mathbf{R})$ . To prove (6.10) we let  $s=0$  and take  $U=W'$ , such that  $U'=W''=V$ .

To prove (6.11), we start by computing the second derivative in (6.12). This time we take  $U=W$ . The computation is valid in the quadratic form sense on  $\mathcal{D}(H^2) \times \mathcal{S}(\mathbf{R})$ .

$$\begin{aligned} \frac{d^2}{ds^2} e^{isH} W e^{-isH_0} &= \frac{d}{ds} e^{isH} (iVW + 2W'p_1 - iW'') e^{-isH_0} \\ &= e^{isH} (iV(iVW + 2W'p_1 - iW'')) \\ &\quad + i[H_0, iVW + 2W'p_1 - iW''] e^{-isH_0} \\ &= e^{isH} (4W''p_1^2 - 4iW^{(3)}p_1 + 2iVW'p_1 + 2i(VW)'p_1 \\ &\quad + (VW)'' - V^2W + VW'' - W^{(4)} - 2W') e^{-isH_0}. \end{aligned}$$

To repeat the argument from the first part we must know that the coefficients to  $p_1$  are of the form  $U'$ . This is obvious in all cases except for  $2VW'$ . However, if we recall  $V=W''$  by assumption, then we see  $2VW'=2W''W'=((W')^2)'$ , and the argument can be repeated. We give the details. Setting  $s=0$  we find the identity.

$$\begin{aligned} (H+i)^{-2} 4W''p_1^2 (H_0+i)^{-2} \\ = (H+i)^{-2} ((HW - WH_0)H_0 - H(HW - WH_0))(H_0+i)^{-2} \\ + (H+i)^{-2} (4iW^{(3)}p_1 - i((W')^2)'p_1 - 2i(VW)'p_1)(H_0+i)^{-2} \\ + (H+i)^{-2} (- (VW)'' + V^2W - VW'' + W^{(4)} + 2W')(H_0+i)^{-2}. \end{aligned}$$

By the first part of the lemma all terms on the right hand side extend to bounded operators on  $L^2(\mathbf{R})$ . Hence the result (6.11) follows.

*Remark 6.6.* The proof of this lemma is somewhat subtle, due to a domain problem. We have  $\mathcal{D}(H) = \mathcal{D}(H_0)$ , since  $V$  is bounded, but in general  $\mathcal{D}(H^2) \neq \mathcal{D}(H_0^2)$ , as can be seen by direct computation in the case of  $V(x_1) = \sin(x_1)$ . Thus it is important to prove (6.11) directly with the factor  $(H+i)^{-2}$  on the left.

**Lemma 6.7.** Suppose  $V = W''$ ,  $W \in \mathcal{B}^4(\mathbf{R})$ . Then the operator  $(H + i)^{-1} V (H_0 + i)^{-1}$  is compact on  $L^2(\mathbf{R})$ .

*Proof.* Using (6.2) it is easy to compute commutators between  $p_1$  and a function of  $H_0$ . In particular, we find, using (6.10) that the operator  $(H + i)^{-1} V (H_0^2 + 1)^{-1/2} (p_1^2 + 1)^{1/2}$  extends to a bounded operator on  $L^2(\mathbf{R})$ . Using complex interpolation we conclude that

$$(H + i)^{-1} V (H_0^2 + 1)^{-\delta/2} (p_1^2 + 1)^{\delta/2}$$

is bounded on  $L^2(\mathbf{R})$  for all  $\delta$ ,  $0 \leq \delta \leq 1$ . We now write

$$\begin{aligned} (H + i)^{-1} V (H_0 + i)^{-1} &= ((H + i)^{-1} V (H_0^2 + 1)^{-1/4} (p_1^2 + 1)^{1/4}) \\ &\quad \cdot ((p_1^2 + 1)^{-1/4} (H_0^2 + 1)^{-1/4}) \\ &\quad \cdot ((H_0^2 + 1)^{1/2} (H_0 + i)^{-1}). \end{aligned}$$

The first term is bounded by the first result ( $\delta = 1/2$ ). The third term is bounded trivially. The second term is a bounded operator, which by (6.2) is unitarily equivalent to  $(p_1^2 + 1)^{-1/4} (x_1^2 + 1)^{-1/4}$ . This operator is compact (see for example [40]).

We summarize the result in the following theorem:

**Theorem 6.8.** Let  $V$  be a realvalued function on  $\mathbf{R}$  such that for some  $W \in \mathcal{B}^4(\mathbf{R})$  we have  $V(x_1) = W''(x_1)$ . Then  $V$  is a short range perturbation of  $H_0 = p_1^2 + x_1$  with respect to  $A = -p_1$ . The wave operators  $W_{\pm}(H_0 + V, H_0)$  exist and are unitary.

*Proof.* The results follows from the abstract short range scattering theory in §4 and Lemmas 6.5 and 6.7, except the result that  $\sigma_{pp}(H) = \emptyset$ . This is a consequence of well-known standard results on ordinary differential equations, see for example [10].

**Remark 6.9.** It is possible to extend the result to include a potential of the type used in Theorem 6.4. It requires some computations and will not be done here. See [20] for the details. Note that for potentials with local singularities we only get that  $\sigma_{pp}(H)$  is discrete in  $\mathbf{R}$ .

Let us give some simple examples of potentials satisfying the assumptions in Theorem 6.8. Let  $V$  be a realvalued function on  $\mathbf{R}$ , which is periodic with period  $P$ ,  $V \in C^2(\mathbf{R})$ , and with  $\int_0^P V(x_1) dx_1 = 0$ . Define

$$W(x_1) = \int_0^{x_1} F(y) dy - c_0 x_1,$$

where

$$F(y) = \int_0^y V(\eta) d\eta, \quad \text{and } c_0 = P^{-1} \int_0^P F(y) dy.$$

Then it is easy to check that all the conditions are satisfied. We can also take a finite sum of such periodic potentials with arbitrary periods.



Assume that  $\mu$  is a complex Borel measure on  $\mathbf{R}$ , with the property

$$\int_{-\infty}^{\infty} (\omega^2 + \omega^{-2}) d|\mu|(\omega) < \infty.$$

Assume the function

$$V(x_1) = \int_{-\infty}^{\infty} e^{-ix_1\omega} d\mu(\omega)$$

is realvalued. Then all conditions are satisfied, if we take

$$W(x_1) = \int_{-\infty}^{\infty} (-\omega^{-2}) e^{-ix_1\omega} d\mu(\omega).$$

As a particular case we can take ( $c_n$  and  $\omega_n$  real)

$$V(x_1) = \sum_{n=1}^{\infty} c_n \sin(\omega_n x_1),$$

if we impose the condition

$$\sum_{n=1}^{\infty} |c_n|(\omega_n^2 + \omega_n^{-2}) < \infty.$$

### 6.3 One-dimensional classical scattering

We now look at the classical scattering theory for the one-dimensional Stark Hamiltonian. At the same time we introduce some of the concepts in scattering theory for a system in classical mechanics. First we discuss results analogous to those in subsection 6.2. Then we look at the one-dimensional problem with potentials of the type discussed in subsection 6.1. One consequence of these computations is that we will discover a discrepancy between classical and quantum scattering for the decaying potentials. This discrepancy will be discussed in the following sections.

We denote by  $C^{2,1}(\mathbf{R})$  the functions  $U \in C^2(\mathbf{R})$  such that the second derivative  $U''$  is Lipschitz-continuous on  $\mathbf{R}$ . In this section we work exclusively in the one-dimensional case. To simplify the notation we write  $(x, p)$  for the variables in the phase space  $\mathbf{R}^2$  instead of  $(x_1, p_1)$ . Note also that here we take the mass of the particle to be equal to 1, whereas in the quantum case we took mass  $\frac{1}{2}$ .

The free classical Stark Hamiltonian is the function

$$H_0(x, p) = \frac{1}{2}p^2 + x, \quad (x, p) \in \mathbf{R}^2.$$

Newton's equation associated with this free Hamiltonian is

$$\ddot{x}(t) = -1, \tag{6.13a}$$

$$\dot{x}(0) = \xi \tag{6.13b}$$

Here we write  $\dot{x}(t)$  and  $\ddot{x}(t)$  for the first and second derivatives with respect to  $t$ . The solution to (6.13) is

$$y(t) = -\frac{1}{2}t^2 + tv + \xi. \quad (6.14)$$

*Assumption 6.10.* Let  $V$  be a realvalued function such that for some  $U \in C^{2,1}(\mathbf{R}) \cap \mathcal{B}^2(\mathbf{R})$  we have  $V(x) = U'(x)$ .

Under this assumption we can consider the full Hamiltonian

$$H(x, p) = \frac{1}{2}p^2 + x + V(x)$$

and the associated Newton's equation

$$\ddot{x}(t) = -1 - V'(x), \quad (6.15a)$$

$$x(0) = x_0, \quad (6.15b)$$

$$\dot{x}(0) = v_0. \quad (6.15c)$$

Under Assumption 6.10 it is well known that solutions to (6.15) exist globally in time. We introduce the following definitions, classifying the solutions by the initial data:

#### DEFINITION 6.11

(i) A solution  $x(t)$  to (6.15) is called a bound state, if

$$\sup_{t \in \mathbf{R}} (|x(t)| + |\dot{x}(t)|) < \infty.$$

In that case we write  $(x_0, v_0) \in \mathcal{M}_{\text{bound}}$ .

(ii) A solution  $x(t)$  to (6.15) is called a scattering state, if there exist  $\xi^\pm$  and  $v^\pm$  such that with

$$y^\pm(t) = -\frac{1}{2}t^2 + tv^\pm + \xi^\pm$$

we have

$$\lim_{t \rightarrow \pm \infty} (|x(t) - y^\pm(t)| + |\dot{x}(t) - \dot{y}^\pm(t)|) = 0.$$

In that case we write  $(x_0, v_0) \in \mathcal{M}_{\text{scat}}$ .

#### DEFINITION 6.12

(i) Let  $(\xi, v) \in \mathbf{R}^2$  be arbitrary. Let  $y(t)$  be given by (6.14). Assume there exist two solutions  $x^\pm$  to (6.15) such that

$$\lim_{t \rightarrow \pm \infty} (|x^\pm(t) - y(t)| + |\dot{x}^\pm(t) - \dot{y}(t)|) = 0.$$

Then the classical wave operators exist and are given by

$$\Omega^\pm(\xi, v) = (x^\pm(0), \dot{x}^\pm(0)). \quad (6.16)$$

(ii) Assume there exists a measurable set  $\mathcal{M}_0 \subset \mathbf{R}^2$  with Lebesgue measure zero, such that

$$\mathbf{R}^2 = \mathcal{M}_{\text{bound}} \cup \mathcal{M}_{\text{scat}} \cup \mathcal{M}_0 \quad (\text{disjoint union}).$$

Then the classical wave operators are said to be asymptotically complete.

*Remark 6.13.* According to this definition the classical wave operators can be complete and not exist at the same time. It is convenient to separate the questions of existence and completeness in this manner.

**Theorem 6.14.** *Let  $V$  satisfy Assumption 6.10. Then the classical wave operators for (6.15) are asymptotically complete.*

*Proof.* We have defined the set  $\mathcal{M}_{\text{scat}}$  as consisting of all initial data leading to scattering states and  $\mathcal{M}_{\text{bound}}$  as those leading to bound states. We also need the sets

$$\mathcal{N}_{\pm} = \{(x_0, v_0) | \text{The solution } x(t) \text{ to (6.15) satisfies } \liminf_{t \rightarrow \pm \infty} x(t) > -\infty\}.$$

An argument due to Littlewood [31] shows that  $\mathcal{N}_+ \setminus \mathcal{N}_-$  and  $\mathcal{N}_- \setminus \mathcal{N}_+$  both have Lebesgue measure zero. See for example [38, Section XI.2] for these computations. We define  $\mathcal{M}_0 = \mathcal{N}_+ \setminus \mathcal{N}_- \cup \mathcal{N}_- \setminus \mathcal{N}_+$ . To complete the proof we must show

$$\mathbf{R}^2 \setminus (\mathcal{M}_0 \cup \mathcal{M}_{\text{bound}}) = \mathcal{M}_{\text{scat}}.$$

We give the details in the case  $t \rightarrow \infty$ . The other case is treated analogously.

Let  $x(t)$  be a solution to (6.15) which is not bounded, i.e.

$$\limsup_{t \rightarrow \infty} (|x(t)| + |\dot{x}(t)|) = +\infty.$$

We have energy conservation, such that

$$E = \frac{1}{2} \dot{x}(t)^2 + x(t) + V(x(t)), \quad t \in \mathbf{R}, \quad (6.17)$$

is constant. Since  $V$  is a bounded function, we get

$$\liminf_{t \rightarrow +\infty} x(t) = -\infty,$$

and

$$\limsup_{t \rightarrow +\infty} |\dot{x}(t)| = +\infty.$$

Using (6.17) once more, we see that there exists  $t_1 > 0$  with  $\dot{x}(t) < 0$  for all  $t \geq t_1$ . Thus

$$\lim_{t \rightarrow +\infty} x(t) = -\infty,$$

and now (6.17) implies

$$\lim_{t \rightarrow +\infty} \dot{x}(t) = \infty$$

Assumption 6.10 implies the existence of  $c_1$  such that

$$|V(y)(1 + V'(y))| \leq c_1 \quad \text{for all } y \in \mathbf{R}.$$

Now fix  $t_0 \geq t_1$  such that

$$\left| \frac{V(x(t))(1 + V'(x(t)))}{\dot{x}(t)^2} \right| \leq \frac{1}{2} \quad \text{for all } t \geq t_0.$$

Take  $t \geq t_0$  and integrate the equation (6.15a) from  $t_0$  to  $t$  to get (using integration by parts)

$$\begin{aligned} \dot{x}(t) - \dot{x}(t_0) &= -(t - t_0) - \int_{t_0}^t V'(x(s))\dot{x}(s)(\dot{x}(s))^{-1} ds \\ &= -(t - t_0) - V(x(t))(\dot{x}(t))^{-1} + V(x(t_0))(\dot{x}(t_0))^{-1} \\ &\quad - \int_{t_0}^t V(x(s))\ddot{x}(s)(\dot{x}(s))^{-2} ds. \end{aligned}$$

We choose a larger  $t_0$  such that the last term is bounded by  $(t - t_0)/2$ . This leads to the estimate

$$\dot{x}(t) \leq -\frac{1}{2}t + c, \quad t \geq t_0.$$

Using this information in the above computation we can now conclude that the limit  $\lim_{t \rightarrow +\infty} (\dot{x}(t) + t)$  exists. Explicitly, using (6.15a) to eliminate  $\ddot{x}(s)$ ,

$$\begin{aligned} \dot{x}(t) + t &= \dot{x}(t_0) + t_0 + V(x(t_0))(\dot{x}(t_0))^{-1} - \int_{t_0}^{\infty} V(x(s))(1 + V'(x(s)))(\dot{x}(s))^{-2} ds \\ &\quad - V(x(t))(\dot{x}(t))^{-1} + \int_t^{\infty} V(x(s))(1 + V'(x(s)))(\dot{x}(s))^{-2} ds \quad (6.18) \end{aligned}$$

Thus the constant term in (6.18) determines  $v_0 = \lim_{t \rightarrow +\infty} (\dot{x}(t) + t)$ . We integrate this equation once more to get

$$\begin{aligned} x(t) + \frac{1}{2}t^2 &= x(t_0) + \frac{1}{2}t_0^2 + v_0(t - t_0) - \int_{t_0}^t V(x(s))(\dot{x}(s))^{-1} ds \\ &\quad + \int_{t_0}^t \int_s^{\infty} V(x(\tau))(1 + V'(x(\tau)))(\dot{x}(\tau))^{-2} d\tau ds. \quad (6.19) \end{aligned}$$

We use integration by parts in the first integral in (6.19), also using  $V = U'$

$$\begin{aligned} \int_{t_0}^t V(x(s))(\dot{x}(s))^{-1} ds &= \int_{t_0}^t U'(x(s))\dot{x}(s)(\dot{x}(s))^{-2} ds = U(x(t))(\dot{x}(t))^{-2} \\ &\quad - U(x(t_0))(\dot{x}(t_0))^{-2} + 2 \int_{t_0}^t U(x(s))\ddot{x}(s)(\dot{x}(s))^{-3} ds. \end{aligned}$$

We see that this term tends to a constant as  $t \rightarrow +\infty$ . We now rewrite the inner

integral in the last term of (6.19)

$$\begin{aligned} \int_s^\infty V(x(\tau))(1 + V'(x(\tau)))\dot{x}(\tau)(\dot{x}(\tau))^{-3}d\tau &= (U(x(s)) + \frac{1}{2}V(x(s))^2)(\dot{x}(s))^{-3} \\ &+ 3 \int_s^\infty (U(x(\tau)) + \frac{1}{2}V(x(\tau))^2)\ddot{x}(\tau)(\dot{x}(\tau))^{-4}d\tau. \end{aligned}$$

The last integral is of order  $O(s^{-3})$  as  $s \rightarrow \infty$ , so the double integral in (6.19) has a limit as  $t \rightarrow \infty$ . These computations prove the existence of

$$\lim_{t \rightarrow +\infty} (x(t) + \frac{1}{2}t^2 - tv_0) = x_0.$$

We have shown that a state, which is not bound, is asymptotic to a free state as  $t \rightarrow +\infty$ . This concludes the proof of the theorem.

**Remark 6.15.** For potentials  $V(x) = O(|x|^{-1/2-\varepsilon})$  as  $x \rightarrow -\infty$  the classical scattering theory was discussed in [4].

We will now turn to the question of existence of the classical wave operators. We have the following result:

**Theorem 6.16.** *Let  $V$  satisfy Assumption 6.10. Let  $(\xi, v) \in \mathbf{R}^2$  and let  $y(t) = -\frac{1}{2}t^2 + tv + \xi$ . Then there exist two solutions  $x^\pm$  to (6.15) such that*

$$\lim_{t \rightarrow \pm\infty} (|x^\pm(t) - y(t)| + |\dot{x}^\pm(t) - \dot{y}(t)|) = 0.$$

As a consequence of Theorems 6.14 and 6.16 we get

**COROLLARY 6.17.**

*The classical wave operators  $\Omega^\pm$ , defined in (6.16), are bijections from  $\mathbf{R}^2$  onto  $\mathcal{M}_{\text{scat}}$ .*

We will not give the proof of Theorem 6.16 here. See Remark 6.25 for some comments.

Next we look at the question of existence and completeness of the classical wave operators in the case where the potential decays slowly in the direction of the electric field. We start with two assumptions on the potential.

**Assumption 6.18.** Let  $V \in \mathcal{B}^1(\mathbf{R})$  be a realvalued function such that  $V'$  is Lipschitz continuous. Assume there exist  $C > 0$ ,  $k \geq 1$ ,  $\alpha > 1$ , and  $x_0 < 0$ , such that

$$|V(x)| \leq C\varphi_\alpha(x) \quad \text{for all } x \leq x_0. \quad (6.20)$$

Here

$$\begin{aligned} \varphi_\alpha(x) = \prod_{j=1}^{k-1} \underbrace{(1 + \log(1 + \log(\cdots(1 + \log(1 + |x|))\cdots)))^{-1}}_{j\text{-times}} \\ \times \underbrace{(1 + \log(1 + \log(\cdots(1 + \log(1 + |x|))\cdots)))^{-\alpha}}_{k\text{-times}} \end{aligned} \quad (6.21)$$

with the convention  $\prod_{j=1}^0(\cdots) = 1$ .

**Assumption 6.19.**  $V$  is a realvalued continuous function on  $\mathbf{R}$ . There exist  $\beta > 1$ ,  $\rho_0 > 1$ , and  $C > 0$ , such that for all  $\rho \geq \rho_0$ ,  $|x_1| \geq \rho$ ,  $|x_2| \geq \rho$ , we have

$$|V(x_1) - V(x_2)| \leq C\varphi_\beta(\rho)|x_1 - x_2|. \quad (6.22)$$

Here  $\varphi_\beta$  is a function of the type (6.21) for some  $k \geq 1$ .

We have stated these rather complicated assumptions because they lead to optimal results. In examples one can think of the typical case  $V \in \mathcal{B}^1(\mathbf{R})$ ,  $V'$  Lipschitz continuous, and

$$|V(x)| + |V'(x)| \leq C(1 + \log|x|)^{-\alpha}, \quad \alpha > 1,$$

for all  $x < -2$ .

Functions of the type (6.21) have the property that the two integrals below are finite for sufficiently large  $t$

$$\int_t^\infty s^{-1} \varphi_\alpha(s) ds < \infty \quad (6.23)$$

$$\int_t^\infty \int_\tau^\infty s^{-2} \varphi_\alpha(s) ds d\tau < \infty. \quad (6.24)$$

**Theorem 6.20.** *Let  $V$  satisfy Assumption 6.18. Then the classical wave operators for (6.15) are asymptotically complete.*

*Proof.* The proof is similar to the proof of Theorem 6.14. We proceed as before until we get to the existence of

$$v = \lim_{t \rightarrow \infty} (\dot{x}(t) + t).$$

Then we have that (6.18) and (6.19) still hold. Now using the properties (6.23) and (6.24) in (6.18) and (6.19) we see that all integrals are absolutely convergent, and we can directly conclude the existence of

$$\xi = \lim_{t \rightarrow \infty} (x(t) + \frac{1}{2}t^2 - vt)$$

which completes the proof.

Next we state and prove the result on the existence of the classical wave operators.

**Theorem 6.21.** *Let  $V$  satisfy Assumptions 6.18 and 6.19. Let  $(\xi, v) \in \mathbf{R}^2$  and let  $y(t) = -\frac{1}{2}t^2 + tv + \xi$ . Then there exist two solutions  $x^\pm$  to (6.15) such that*

$$\lim_{t \rightarrow \pm \infty} (|x^\pm(t) - y(t)| + |\dot{x}^\pm(t) - \dot{y}(t)|) = 0.$$

*Proof.* The proof is rather long. The idea is to use a fixed point argument at infinity. We consider only the  $+$ -case. The superscript  $+$  is omitted in the sequel. Let  $(\xi, v) \in \mathbf{R}^2$  be fixed and let  $y(t) = -\frac{1}{2}t^2 + tv + \xi$ . We try to find  $x(t)$  in the form  $x(t) = y(t) + u(t)$ .

Then  $u(t)$  must satisfy the equation

$$\ddot{u}(t) = -V'(y(t) + u(t)). \quad (6.25)$$

This differential equation is transformed into a pair of integral equations, as usual. We need an appropriate space to work in. Let  $t_0 \geq 1$  be a parameter which we will fix later. We define

$$\mathcal{B}(t_0) = \{(u, w) | u, w \in C((t_0, \infty), \mathbf{R}), \quad |u(t)| + |w(t)| \rightarrow 0 \text{ as } t \rightarrow \infty, \\ |u(t)| + |w(t)| \leq \frac{1}{2} \text{ for all } t \geq t_0\}. \quad (6.26)$$

We equip  $\mathcal{B}(t_0)$  with the metric

$$d((u_1, w_1), (u_2, w_2)) = \|u_1 - u_2\|_\infty + \|w_1 - w_2\|_\infty.$$

Then  $\mathcal{B}(t_0)$  becomes a complete metric space. Now we define a map by  $J(u, w) = (\tilde{u}, \tilde{w})$ , where we use the computations leading to (6.18) and (6.19) to get the right form of the map.

$$\begin{aligned} \tilde{u}(t) = & \int_t^\infty V(y(s) + u(s))(\dot{y}(s) + w(s))^{-1} ds \\ & - \int_t^\infty \int_\tau^\infty V(y(s) + u(s))(1 + V'(y(s) + u(s)))(\dot{y}(s) + w(s))^{-2} ds d\tau. \end{aligned} \quad (6.27)$$

$$\begin{aligned} \tilde{w}(t) = & -V(y(t) + u(t))(\dot{y}(t) + w(t))^{-1} \\ & + \int_t^\infty V(y(s) + u(s))(1 + V'(y(s) + u(s)))(\dot{y}(s) + w(s))^{-2} ds. \end{aligned} \quad (6.28)$$

Now we need the following lemma:

*Lemma 6.22. There exists  $T_0 > 1$  such that for all  $t_0 \geq T_0$  the map  $J$  is well-defined on  $\mathcal{B}(t_0)$  and maps  $\mathcal{B}(t_0)$  into  $\mathcal{B}(t_0)$ . Furthermore, there exists a constant  $c$ ,  $0 < c < 1$ , such that for all  $(u_1, w_1), (u_2, w_2) \in \mathcal{B}(t_0)$  we have*

$$d(J(u_1, w_1), J(u_2, w_2)) \leq cd((u_1, w_1), (u_2, w_2)),$$

such that  $J$  is a contraction on  $\mathcal{B}(t_0)$ .

*Proof of the lemma.* Recall that we have fixed  $(\xi, v) \in \mathbf{R}^2$ . We can then determine a constant  $s_0$  such that

$$\begin{aligned} |y(s) + u(s)| & \geq Cs^2, \\ |\dot{y}(s) + w(s)| & \geq Cs, \end{aligned} \quad (6.29)$$

for all  $s \geq s_0$  and all  $(u, w) \in \mathcal{B}(s_0)$ . Using Assumption 6.18 and definition (6.27) we get for all  $t \geq t_0 \geq s_0$  the estimate

$$|\tilde{u}(t)| \leq C \int_t^\infty s^{-1} \varphi_\alpha(s) ds + C \int_t^\infty \int_\tau^\infty s^{-2} \varphi_\alpha(s) ds d\tau. \quad (6.30)$$

We see from (6.27) and (6.29) that  $\tilde{u}(t)$  is a well-defined continuous function on  $(t_0, \infty)$ . Furthermore,  $\tilde{u}(t) \rightarrow 0$  as  $t \rightarrow \infty$  by (6.30). If we choose  $s_0$  sufficiently large, then we can get  $|\tilde{u}(t)| \leq \frac{1}{4}$  for all  $t \geq t_0 \geq s_0$ .

For  $\tilde{w}(t)$  we get the estimate

$$|\tilde{w}(t)| \leq Ct^{-1} + C \int_t^\infty s^{-2} \varphi_\alpha(s) ds, \quad (6.31)$$

for all  $t \geq t_0 \geq s_0$ . We can then repeat the arguments to conclude that  $\tilde{w}(t)$  is a well-defined continuous function,  $\tilde{w}(t) \rightarrow 0$  as  $t \rightarrow \infty$ , and  $|\tilde{w}(t)| \leq \frac{1}{4}$  for all  $t \geq t_0 \geq s_0$ , provided we take  $s_0$  sufficiently large.

The Lipschitz-condition (6.22) in Assumption 6.19 and some straightforward computations show that if we take  $T_0$  sufficiently large, then the contraction property holds. Further details are omitted.

*Proof of the theorem continued.* Assume  $t_0 \geq T_0$  from Lemma 6.22. The fixed point theorem for contractions yields the existence of (a unique)  $(u, w) \in \mathcal{B}(t_0)$  with  $J(u, w) = (u, w)$ . We must now verify that  $u \in C^2((t_0, \infty))$  and that  $u$  satisfies (6.25). We write out the equations (6.27) and (6.28) for the fixed point:

$$\begin{aligned} u(t) = & \int_t^\infty V(y(s) + u(s))(\dot{y}(s) + w(s))^{-1} ds \\ & - \int_t^\infty \int_\tau^\infty V(y(s) + u(s))(1 + V'(y(s) + u(s)))(\dot{y}(s) + w(s))^{-2} ds d\tau. \end{aligned} \quad (6.32)$$

$$\begin{aligned} w(t) = & -V(y(t) + u(t))(\dot{y}(t) + w(t))^{-1} \\ & + \int_t^\infty V(y(s) + u(s))(1 + V'(y(s) + u(s)))(\dot{y}(s) + w(s))^{-2} ds. \end{aligned} \quad (6.33)$$

It follows immediately that  $u \in C^1((t_0, \infty))$  and

$$\dot{u}(t) = w(t). \quad (6.34)$$

We introduce the abbreviations

$$f(t) = -V(y(t) + u(t))$$

and

$$g(t) = \int_t^\infty V(y(s) + u(s))(1 + V'(y(s) + u(s)))(\dot{y}(s) + w(s))^{-2} ds.$$

Then  $f, g \in C^1((t_0, \infty))$  and the equation (6.33) is written as

$$w(t) = f(t)(-t + v + w(t))^{-1} + g(t)$$

or

$$w(t)^2 + (-t + v - g(t))w(t) + (t - v)g(t) - f(t) = 0. \quad (6.35)$$

This quadratic equation in  $w$  has the discriminant

$$D(t) = (-t + v - g(t))^2 - 4((t - v)g(t) - f(t)).$$



Since both  $f$  and  $g$  are bounded functions, we can find  $t_1 \geq t_0$  such that for all  $t \geq t_1$  we have  $D(t) \geq \frac{1}{2}t^2$ . It follows that either solution to the quadratic equation (6.35) belong to  $C^1((t_1, \infty))$ . Thus we get that  $u \in C^2((t_1, \infty))$  and  $\ddot{u} = \dot{w}$ . Differentiate both sides of (6.33) and substitute  $\dot{u} = w$  to get

$$\begin{aligned}\ddot{u}(t) = \dot{w}(t) = & -V'(y(t) + u(t)) + V(y(t) + u(t))(-1 + \ddot{u}(t))(\dot{y}(t) + \dot{u}(t))^{-2} \\ & + V(y(t) + u(t))(1 + V'(y(t) + u(t)))(\dot{y}(t) + \dot{u}(t))^{-2}.\end{aligned}$$

Rewrite as

$$(\ddot{u}(t) + V'(y(t) + u(t)))(1 - V(y(t) + u(t))(\dot{y}(t) + \dot{u}(t))^{-2}) = 0.$$

Now fix  $t_2 \geq t_1$  such that

$$|V(y(t) + u(t))(\dot{y}(t) + \dot{u}(t))^{-2}| \leq \frac{1}{2}$$

for all  $t \geq t_2$ . We conclude that for  $t \geq t_2$

$$\ddot{u}(t) = -V'(y(t) + u(t))$$

which proves (6.25). By uniqueness and global existence of solutions to (6.25) this result holds for all  $t \in \mathbf{R}$ . This concludes the proof of the theorem.

We have the same corollary as before, this time to Theorems 6.20 and 6.21.

#### COROLLARY 6.23.

Let  $V$  satisfy Assumptions 6.18 and 6.19. Then the classical wave operators  $\Omega^\pm$ , defined in (6.16), are bijections from  $\mathbf{R}^2$  onto  $\mathcal{M}_{\text{scat}}$ .

*Remark 6.24.* We have used the complicated function  $\varphi_\alpha$  in Assumption 6.18, since it gives an almost optimal condition for asymptotic completeness. This can be seen as follows: Take a potential  $V \in C^2(\mathbf{R})$ , such that  $V(x) = \varphi_1(x)$ ,  $x < -2$ ,  $V(x) = 0$ ,  $x > 0$ . Here  $\varphi_1$  is any function given by (6.21) with  $\alpha = 1$ . Using the computations in the proof of Theorems 6.14 and 6.20 it is easy to see that the limit

$$\lim_{t \rightarrow \infty} (x(t) + \frac{1}{2} - tv)$$

does not exist.

Analyzing the proof further it is clear that we only use the properties (6.23) and (6.24), together with monotonicity of  $\varphi_\alpha$ .

*Remark 6.25.* Let us explain how to prove Theorem 6.16. We take the space  $\mathcal{B}(t_0)$  defined in (6.26), but now the operator  $J$  is given by  $J(u, w) = (\tilde{u}, \tilde{w})$ , where

$$\begin{aligned}\tilde{u}(t) = & -U(y(t) + u(t))(\dot{y}(t) + w(t))^{-2} \\ & - \int_t^\infty \{3U(y(s) + u(s)) + 2U(y(s) + u(s)) V'(y(s) + u(s)) \\ & + \frac{1}{2} V(y(s) + u(s))^2\} (\dot{y}(s) + w(s))^{-3} ds\end{aligned}$$

$$\begin{aligned}
& + 3 \int_t^\infty \int_\tau^\infty (U(y(s) + u(s)) + \tfrac{1}{2} V(y(s) + u(s))^2) \\
& \cdot (1 + V'(y(s) + u(s))) (\dot{y}(s) + w(s))^{-4} ds d\tau, \\
\tilde{w}(t) = & -V(y(t) + u(t)) (\dot{y}(t) + w(t))^{-1} \\
& + \int_t^\infty V(y(s) + u(s)) (1 + V'(y(s) + u(s))) (\dot{y}(s) + w(s))^{-2} ds.
\end{aligned}$$

Given these definitions the idea is again to show that for  $t_0$  sufficiently large  $J$  is a contraction on  $\mathcal{B}(t_0)$ , and then to show that  $x(t) = y(t) + u(t)$ , with  $u(t)$  first component of the fixed point, solves Newton's equation. The computations are quite tedious.

#### 6.4 The discrepancy

Comparing the results in § 6.2 and § 6.3 we observe a remarkable discrepancy. We see that in the classical case the wave operators exist and are complete for potentials with decay essentially

$$|V(x)| + |V'(x)| \leq C(1 + \log|x|)^{-\alpha}, \quad \alpha > 1.$$

In the quantum case we had to have either oscillation (see Theorem 6.8) or decay  $V(x) = O(|x|^{-1/2-\varepsilon})$  as  $x \rightarrow -\infty$ . The decay result cannot be improved, since a result by Ozawa [34] shows that for homogeneous potentials with the behavior  $|x|^{-1/2}$  the wave operators fail to exist. This result holds in any dimension. A general non-existence result which covers this case is given in Theorem 8.2. We will put all results together and resolve the discrepancy in § 10.

### 7. Stark Hamiltonians in the moving frame picture

To continue our discussion of Stark Hamiltonians, and in particular the discrepancy mentioned in § 6.4, it is convenient to introduce the moving frame picture. We will limit the discussion to the one-dimensional case (the extension to higher dimensions is trivial), and continue writing  $x$  instead of  $x_1$ . We also use the notation  $p = -id/dx$ .

Let  $\Phi \in L^\infty(\mathbb{R})$  be a realvalued function. We consider in  $L^2(\mathbb{R})$  the operators

$$H_0 = \tfrac{1}{2}p^2 + x, \tag{7.1a}$$

$$H = \tfrac{1}{2}p^2 + x + \Phi(x), \tag{7.1b}$$

and the operators

$$K_0 = \tfrac{1}{2}p^2, \tag{7.2a}$$

$$K(t) = \tfrac{1}{2}p^2 + V(t, x), \tag{7.2b}$$

$$V(t, x) = \Phi(x - \tfrac{1}{2}t^2), \quad t \in \mathbb{R}. \tag{7.2c}$$

The operators in (7.1) are the usual free and full Stark Hamiltonians, and the operators

in (7.2) are the same pair in the moving frame picture. We introduce the operator

$$T(t) = e^{-it^3/6} e^{-itx} e^{it^2 p/2}, \quad t \in \mathbf{R} \quad (7.3)$$

on  $L^2(\mathbf{R})$ , and recall the formula from [4] (see also (6.2))

$$e^{-itH_0} = T(t) e^{-itK_0}. \quad (7.4)$$

Now with the correspondence

$$u(t) = T(t)v(t) \quad (7.5)$$

the problems

$$i\partial_t u = H_0 u, \quad (7.6a)$$

$$i\partial_t u = H u, \quad (7.6b)$$

are equivalent to the problems

$$i\partial_t v = K_0 v, \quad (7.7a)$$

$$i\partial_t v = K(t)v. \quad (7.7b)$$

For (7.6a) and (7.7a) this is a consequence of (7.4). In the other case computations are needed. First we must show that (7.7b) with  $K(t)$  from (7.2b) and (7.2c) has a unique unitary propagator for a suitable class of functions  $\Phi$ . If we impose enough conditions, this is a consequence of general results on Schrödinger operators with time-dependent potentials. Here we will use a more direct approach. Assume  $\Phi \in L^\infty(\mathbf{R})$  is a real valued function. Then we define the propagator by

$$U(t, s) = T(t)^{-1} e^{-i(t-s)H} T(s), \quad t, s \in \mathbf{R}. \quad (7.8)$$

The operators  $U(t, s)$  are unitary and satisfy

$$U(t, t) = 1, \quad t \in \mathbf{R}, \quad (7.9)$$

$$U(t, s)U(s, r) = U(t, r), \quad t, s, r \in \mathbf{R}, \quad (7.10)$$

trivially. Formally we have

$$i\partial_t U(t, s) = K(t)U(t, s), \quad (7.11)$$

$$i\partial_s U(t, s) = -U(t, s)K(s). \quad (7.12)$$

To verify these relations we introduce the space  $\mathcal{H} = \mathcal{D}(p^2) \cap \mathcal{D}(x)$  with the norm

$$\|\varphi\|_{\mathcal{H}} = (\|\varphi\|^2 + \|p^2 \varphi\|^2 + \|x\varphi\|^2)^{1/2}.$$

We have the following result:

*Lemma 7.1. Assume  $\Phi \in \mathcal{B}^1(\mathbf{R})$ . Then we have*

(i)  $U(t, s)$  maps  $\mathcal{H}$  into  $\mathcal{H}$ , and  $U(t, s)$  is strongly continuous in  $t, s$  on  $\mathcal{H}$ .

$$i\partial_s U(t, s)\varphi = -U(t, s)K(s)\varphi.$$

The proof consists in straightforward computations based on the definition (7.8) and is omitted.

Now  $U(t, s)$  has enough properties that we can proceed to the scattering theory. Using (7.4) and (7.8) we see

$$\begin{aligned} W_+(H, H_0) &= s\text{-}\lim_{t \rightarrow \infty} e^{itH} e^{-itH_0} \\ &= s\text{-}\lim_{t \rightarrow \infty} U(0, t) e^{-itK_0}. \end{aligned}$$

Thus we can study the scattering problem for (7.2) instead of (7.1). This problem will be considered in some generality in § 8 and § 9 before we return to Stark Hamiltonians in § 10.

Let us briefly note that in the general case with the electric field given by  $F \in \mathbf{R}^d$ ,  $F \neq 0$ , the correct Hamiltonian with time-dependent potential is

$$K(t) = -\frac{1}{2}\Delta + \Phi(x - \frac{1}{2}t^2 F).$$

## 8. Existence and non-existence of wave operators for time-dependent potentials

We give a result on both existence and non-existence of the wave operators in the general case where we have time-dependent potentials. We apply the non-existence results to Stark Hamiltonians. The results in this section were first obtained in [24], where further results and applications can be found.

In this section we continue using the notation from § 7, writing  $K_0 = -\frac{1}{2}\Delta$  and

$$K(t) = K_0 + V(t, \cdot).$$

The problem

$$i\partial_t \varphi = K(t)\varphi, \quad \varphi(s) = \varphi_0, \tag{8.1}$$

is solved by  $\varphi(t) = U(t, s)\varphi_0$ , where  $U(t, s)$  is the unitary propagator associated with the problem, provided it exists. We impose

*Assumption 8.1.*  $V: \mathbf{R} \times \mathbf{R}^d \rightarrow \mathbf{R}$  is a measurable function such that there is a unique unitary propagator associated with the problem (8.1).

Our proofs are based on a configuration space factorization of the free propagator  $U_0(t) = \exp(-itK_0)$  due to Dollard. We have

$$U_0(t) = M(t)D(t)FM(t), \tag{8.2}$$

where  $M(t) = \exp(ix^2/(2t))$  is a multiplication operator,  $D(t)$  is the modified dilation

and  $F\varphi = \hat{\varphi}$  denotes the Fourier transform of  $\varphi$ . The proof is based on the operator relation

$$(i\partial_t + \frac{1}{2}\Delta)M(t)D(t) = M(t)D(t)\left(i\partial_t + \frac{1}{2t^2}\Delta\right), \quad (8.3)$$

which is easily verified. We introduce the wave operators

$$W_{\pm}(0) = s\text{-}\lim_{t \rightarrow \pm\infty} U(0, t)U_0(t).$$

We let  $\chi_K$  denote multiplication by the characteristic function of the set  $K \subset \mathbf{R}^d$ . The main result will be stated only in the  $+$ -case.

**Theorem 8.2.** *Let  $V$  satisfy Assumption 8.1. Assume that there exist  $t_0 > 0$ ,  $\mathcal{W} \in L^2_{\text{loc}}(\mathbf{R}^d \setminus N)$ ,  $N$  a closed set of measure zero,  $m \geq 0$ , and  $\theta \in C([t_0, \infty); \mathbf{R}_+)$  with  $\int_{t_0}^{\infty} \theta(\tau) d\tau = +\infty$ . Furthermore, assume that for each compact  $K \subset \mathbf{R}^d \setminus N$  there exist  $\rho_K \in L^1([t_0, \infty))$ ,  $\rho_K \geq 0$ , and  $\sigma_K \in C([t_0, \infty); \mathbf{R}_+)$ ,  $\sigma_K(t) \rightarrow 0$  as  $t \rightarrow \infty$ , such that for all  $t \geq t_0$*

$$\|\chi_K \{D(t)^{-1}V(t, \cdot)D(t) - \theta(t)\mathcal{W}\}(1 - \Delta)^{-m}\|_{\mathcal{B}(L^2(\mathbf{R}^d))} \leq \sigma_K(t)\theta(t) + \rho_K(t). \quad (8.4)$$

Then the following two results hold:

- (i) Let  $\varphi \in L^2(\mathbf{R}^d)$ . Assume  $\varphi_+ = \lim_{t \rightarrow \infty} U(0, t)U_0(t)\varphi$  exists. Then  $\mathcal{W}\hat{\varphi} = 0$ .
- (ii) Assume  $\sigma_K\theta \in L^1([t_0, \infty))$  for all  $K \subset \mathbf{R}^d \setminus N$  compact. Let  $\varphi \in L^2(\mathbf{R}^d)$  with  $\mathcal{W}\hat{\varphi} = 0$ . Then  $\lim_{t \rightarrow \infty} U(0, t)U_0(t)\varphi$  exists in  $L^2(\mathbf{R}^d)$ .

*Proof.* We first prove (i). Let  $\psi \in \mathcal{S}(\mathbf{R}^d)$ ,  $\hat{\psi} \in C_0^\infty(\mathbf{R}^d \setminus N)$ , and take  $K \subset \mathbf{R}^d \setminus N$  compact,  $K \supseteq \text{supp}(\hat{\psi})$ . We assume  $(\mathcal{W}\hat{\varphi}, \hat{\psi}) \neq 0$  and derive a contradiction. Let  $t > s > t_0$ . By the Cook-Kuroda argument (see § 3.2) we have

$$\begin{aligned} & i(\varphi_+, U(0, t)M(t)D(t)\hat{\psi} - U(0, s)M(s)D(s)\hat{\psi}) \\ &= \int_s^t i \frac{d}{d\tau} (\varphi_+, U(0, \tau)M(\tau)D(\tau)\hat{\psi}) d\tau \\ &= \int_s^t (\varphi_+, U(0, \tau)(i\partial_\tau - K(\tau))M(\tau)D(\tau)\hat{\psi}) d\tau. \end{aligned}$$

Now use (8.3) to get

$$(i\partial_\tau - K(\tau))M(\tau)D(\tau)\hat{\psi} = -V(\tau)M(\tau)D(\tau)\hat{\psi} + M(\tau)D(\tau)\frac{1}{2\tau^2}(\Delta\hat{\psi}).$$

The expression above is rewritten by adding and subtracting some terms. We omit

the argument  $\tau$  in several places to simplify the notation.

$$\begin{aligned}
 & i(\varphi_+, U(0, t)M(t)D(t)\hat{\psi} - U(0, s)M(s)D(s)\hat{\psi}) \\
 &= - \int_s^t (MD\hat{\phi}, MD(D^{-1}V(\tau)D)\hat{\psi})d\tau \\
 &+ \int_s^t (MDF(\varphi - M\varphi), MD(D^{-1}V(\tau)D)\hat{\psi})d\tau \\
 &+ \int_s^t (U_0(\tau)\varphi - U(\tau, 0)\varphi_+, MD(D^{-1}V(\tau)D)\hat{\psi})d\tau \\
 &+ \int_s^t \left( U(\tau, 0)\varphi_+, MD \frac{1}{2\tau^2} \Delta \hat{\psi} \right) d\tau \\
 &= \text{I} + \text{II} + \text{III} + \text{IV}.
 \end{aligned}$$

The four terms are estimated below:

$$\begin{aligned}
 |\text{I}| &= \left| \int_s^t (\hat{\phi}, D^{-1}V(\tau)D\hat{\psi})d\tau \right| \\
 &= \left| \int_s^t \theta(\tau)(\hat{\phi}, \mathcal{W}\hat{\psi})d\tau + \int_s^t (\hat{\phi}, \chi_K(D^{-1}V(\tau)D - \theta(\tau)\mathcal{W})\hat{\psi})d\tau \right| \\
 &\geq |(\hat{\phi}, \mathcal{W}\hat{\psi})| \int_s^t \theta(\tau)d\tau - \int_s^t (\sigma_K(\tau)\theta(\tau) + \rho_K(\tau))d\tau \cdot \|(1 - \Delta)^m \hat{\psi}\| \|\hat{\phi}\|. \\
 |\text{II}| &= \left| \int_s^t (F(\varphi - M\varphi), V(\tau, \tau(\cdot))\hat{\psi})d\tau \right| \\
 &\leq \int_s^t \|\chi_K V(\tau, \tau(\cdot))\hat{\psi}\| \|(1 - M(\tau))\varphi\| d\tau \\
 &\leq \int_s^t \{ \|\chi_K \{D(\tau)^{-1}V(\tau)D(\tau) - \theta(\tau)\mathcal{W}\}(1 - \Delta)^{-m}\|_{\mathcal{B}(L^2(\mathbb{R}^d))} \\
 &\quad \| (1 - \Delta)^m \hat{\psi} \| + \theta(\tau) \|\mathcal{W}\hat{\psi}\| \} \|(1 - M(\tau))\varphi\| d\tau \\
 &\leq \int_s^t \{ (\sigma_K(\tau)\theta(\tau) + \rho_K(\tau)) \|(1 - \Delta)^m \hat{\psi}\| + \theta(\tau) \|\mathcal{W}\hat{\psi}\| \} \\
 &\quad \cdot \|(1 - M(\tau))\varphi\| d\tau.
 \end{aligned}$$

The estimate for the term III is similar to the one for II. We get

$$\begin{aligned}
 |\text{III}| &= \int_s^t \|\chi_K V(\tau, \tau(\cdot))\hat{\psi}\| \|U(\tau)\varphi - U(\tau, 0)\varphi_+\| d\tau \\
 &\leq \int_s^t \{ (\sigma_K(\tau)\theta(\tau) + \rho_K(\tau)) \|(1 - \Delta)^m \hat{\psi}\| + \theta(\tau) \|\mathcal{W}\hat{\psi}\| \} \\
 &\quad \cdot \|U(\tau)\varphi - U(\tau, 0)\varphi_+\| d\tau.
 \end{aligned}$$

Finally we have

$$|IV| \leq \int_s^t \|\Delta \hat{\psi}\| \|\varphi_+\| (2\tau^2)^{-1} d\tau.$$

We now put these estimates together. Given  $\varepsilon > 0$ , we can find  $T(\varepsilon) > t_0$  such that for all  $t > s > T(\varepsilon)$

$$\begin{aligned} 2\|\psi\| \|\varphi_+\| &\geq |( \varphi_+, U(0, t)M(t)D(t)\hat{\psi} - U(0, s)M(s)D(s)\hat{\psi} )| \\ &\geq \{ |(\mathcal{W}\hat{\phi}, \hat{\psi})| - \varepsilon \} \int_s^t \theta(\tau) d\tau - cs^{-1} - C \int_s^\infty \rho_K(\tau) d\tau. \end{aligned}$$

Take  $0 < \varepsilon < |(\mathcal{W}\hat{\phi}, \hat{\psi})|$ . If we let  $t \rightarrow \infty$ , then we get a contradiction. We conclude that

$$(\mathcal{W}\hat{\phi}, \hat{\psi}) = 0.$$

Since  $\hat{\psi}$  varies through a dense subset of  $L^2(\mathbf{R}^d)$ , we get  $\mathcal{W}\hat{\phi} = 0$ . This completes the proof of part (i).

To prove part (ii) it suffices to consider  $\varphi \in \mathcal{S}(\mathbf{R}^d)$  with  $\hat{\phi} \in C_0^\infty(\mathbf{R}^d \setminus N)$  and  $\mathcal{W}\hat{\phi} = 0$ , by a simple approximation argument. Let us fix  $\varphi$  with these properties, and let  $K = \text{supp}(\hat{\phi})$ . We repeat the above computations, using the condition  $\mathcal{W}\hat{\phi} = 0$  to get

$$\begin{aligned} &\| (U(0, t)M(t)D(t)\hat{\phi} - U(0, s)M(s)D(s)\hat{\phi}) \| \\ &= \left\| \int_s^t U(0, \tau)(i\partial_\tau - K(\tau))M(\tau)D(\tau)\hat{\phi} d\tau \right\| \\ &\leq \int_s^t \|D(\tau)^{-1}V(\tau)D(\tau)\hat{\phi}\| d\tau + \frac{1}{2} \int_s^t \tau^{-2} \|\Delta\hat{\phi}\| d\tau \\ &\leq \int_s^t \|\chi_K\{D(\tau)^{-1}V(\tau)D(\tau) - \theta(\tau)\mathcal{W}\}(1-\Delta)^{-m}\|_{\mathcal{B}(L^2(\mathbf{R}^d))} \\ &\quad \cdot \|(1-\Delta)^m\hat{\phi}\| d\tau + \int_s^t \tau^{-2} \|\Delta\hat{\phi}\| d\tau. \end{aligned}$$

Due to the assumptions in (ii) we conclude the existence of

$$\lim_{t \rightarrow \infty} U(0, t)M(t)D(t)F\varphi$$

Since  $(1 - M(t))\varphi \rightarrow 0$  as  $t \rightarrow \infty$ , we get from (8.2) the existence of

$$\lim_{t \rightarrow \infty} U(0, t)U_0(t)\varphi.$$

This concludes the proof of part (ii).

We apply the non-existence part to show that  $O(|x|^{-1/2})$  is the borderline behavior for Stark Hamiltonians. Take  $F \in \mathbf{R}^d$ ,  $F \neq 0$ ,  $\Phi(x) = \lambda|x|^{-\mu}$ ,  $\lambda \neq 0$ , and let  $V(t, x) = \Phi(x - \frac{1}{2}t^2F)$ . We now verify the conditions in Theorem 8.2. Take

$N = \{sF | s \in \mathbf{R}\}$  and let  $K \subset \mathbf{R}^d \setminus N$  be a compact set. We define

$$\sigma_K(t) = |\lambda| \max_{x \in K} ||t^{-1}x - \frac{1}{2}F|^{-\mu} - |\frac{1}{2}F|^{-\mu}|, \quad t \geq 1,$$

$$\theta(t) = t^{-2\mu}, \quad t \geq 1,$$

$$\mathcal{W}(x) = |\lambda| |\frac{1}{2}F|^{-\mu}, \quad x \in \mathbf{R}^d.$$

Thus we have

$$|\chi_K(x)(V(t, tx) - \mathcal{W}(x))| \leq \theta(t)\sigma_K(t)$$

for all  $x \in \mathbf{R}^d$  and all  $t \geq 1$ , which implies the estimate (8.4) with  $m = 0$  and  $\rho_K(x) \equiv 0$ . If  $0 < \mu \leq 1/2$ , all conditions in the theorem are satisfied. In the one-dimensional case we have to take  $0 < \mu < 1/2$  in order to get an operator perturbation of  $H_0$ .

One can use the theorem to prove that the Coulomb potential  $c/|x|$  gives the borderline for the existence of wave operators for perturbations of  $-\Delta$ , if one excludes oscillating potentials. See Propositions 5.4 and 5.5 for some results in the positive direction.

## 9. Existence of modified wave operators

For the Coulomb potential Dollard [9] proposed a modification of the free evolution to take into account the long range nature of the interaction, and to get modified wave operators with the intertwining property such that they could be used in the spectral analysis of the full Hamiltonian. Many researchers worked on generalizing this modification to potentials with decay rate  $V(x) = O(|x|^{-\varepsilon})$ ,  $\varepsilon > 0$ , since the Dollard modifier only works for  $\varepsilon > 1/2$ . See e.g. [16, 17, 38] for discussion of this problem.

In this section we introduce the Dollard-type modified wave operators for time-dependent potentials. We assume a splitting of the potential

$$V(t, x) = V_s(t, x) + V_l(t, x)$$

into a short range and a long range part. We then introduce the function

$$S(t, y) = - \int_0^t V_l(\tau, \tau y) d\tau. \quad (9.1)$$

To simplify the notation, the following convention is used:  $S(t)$  or  $S(t, \cdot)$  denote the operator of multiplication by the function  $S(t, x)$ , and  $S(t, p)$  denotes the operator defined via the functional calculus (or the Fourier transform), where as before  $p = -i\nabla$ . The Dollard-type modified free evolution is given by

$$U_D(t) = \exp(-it\frac{1}{2}p^2 + iS(t, p)). \quad (9.2)$$

We assume that there is a unitary propagator  $U(t, s)$  associated with the Hamiltonian  $K(t) = \frac{1}{2}p^2 + V(t, \cdot)$ . We continue to use the notation of the previous two sections for



perturbations of the Laplacian. The Dollard-type modified wave operator is given by

$$W_{\pm}^D = s\text{-}\lim_{t \rightarrow \pm \infty} U(0, t) U_D(t). \quad (9.3)$$

We give one result on the existence of  $W_{\pm}^D$ .

**Theorem 9.1.** *Let  $V(t, x)$  be a realvalued function such that there is a unique unitary propagator  $U(t, s)$  associated with  $K(t) = -\frac{1}{2}\Delta + V(t, x)$ . Assume we have a decomposition  $V = V_s + V_l$ . Assume there exist a closed set  $N$  of measure zero and a number  $m \geq 0$  such that the following conditions are satisfied:*

(i) *For every compact set  $K \subset \mathbb{R}^d \setminus N$  the operator*

$$\chi_K(\cdot) V_s(t, t(\cdot))(1 - \Delta)^{-m}$$

*is bounded on  $L^2(\mathbb{R}^d)$ , and for some  $t_0 \geq 0$*

$$\int_{|t| \geq t_0} \|\chi_K(\cdot) V_s(t, t(\cdot))(1 - \Delta)^{-m}\|_{\mathcal{B}(L^2(\mathbb{R}^d))} dt < \infty. \quad (9.4)$$

(ii) *For every compact set  $K \subset \mathbb{R}^d \setminus N$  there exists  $t_K > 0$  such that  $\nabla S(t) \in L^\infty(K)$ ,  $\Delta S(t) \in L^\infty(K)$  for all  $t, |t| \geq t_K$ , and such that*

$$\lim_{|t| \rightarrow \infty} |t|^{-1/2} \|\nabla S(t)\|_{L^\infty(K)} = 0, \quad (9.5)$$

$$\int_{|t| \geq t_K} t^{-2} \|\nabla S(t)\|_{L^\infty(K)}^2 dt < \infty, \quad (9.6)$$

$$\int_{|t| \geq t_K} t^{-2} \|\Delta S(t)\|_{L^\infty(K)} dt < \infty. \quad (9.7)$$

*Then the wave operators*

$$W_{\pm}^D = s\text{-}\lim_{t \rightarrow \pm \infty} U(0, t) U_D(t)$$

*exist.*

*Proof.* We consider only  $W_+^D$ . Let  $\varphi \in S(\mathbb{R}^d)$  with  $\hat{\varphi} \in C_0^\infty(\mathbb{R}^d \setminus N)$  and let  $K = \text{supp}(\hat{\varphi})$ . We note the following consequence of our definitions:

$$V(t, tx) + \partial_t S(t) = V_s(t, tx). \quad (9.8)$$

Using the factorization (8.2) and equation (8.3) we get after a simple computation

$$\begin{aligned} & (i\partial_t - K(t))(MD \exp(iS(t)))\hat{\varphi} \\ &= MD \exp(iS(t)) \left\{ -V_s(t, tx)\hat{\varphi} - \frac{1}{2t^2}(\nabla S(t))^2 \hat{\varphi} + \frac{i}{2t^2}(\Delta S(t))\hat{\varphi} \right. \\ & \quad \left. + \frac{1}{2t^2}\Delta \hat{\varphi} + \frac{i}{t^2}\nabla S(t) \cdot \nabla \hat{\varphi} \right\}. \end{aligned} \quad (9.9)$$

and then we use (9.9) to get for  $t > s > t_K$

$$\begin{aligned} & \| U(0, t) M(t) D(t) \exp(iS(t)) \hat{\phi} - U(0, s) M(s) D(s) \exp(iS(s)) \hat{\phi} \| \\ &= \left\| \int_s^t U(0, \tau) (i\partial_\tau - K(\tau)) M(\tau) D(\tau) \exp(iS(\tau)) \hat{\phi} d\tau \right\| \\ &\leq \int_s^t \| V_s(\tau, \tau(\cdot)) \hat{\phi} \| d\tau \end{aligned} \quad (9.10)$$

$$+ \int_s^t \tau^{-2} (\| (\nabla S(\tau))^2 \hat{\phi} \| + \| (\Delta S(\tau)) \hat{\phi} \|) d\tau \quad (9.11)$$

$$+ \int_s^t \tau^{-2} \| \nabla S(\tau) \cdot \nabla \hat{\phi} \| d\tau \quad (9.12)$$

$$+ \int_s^t \tau^{-2} \| \Delta \hat{\phi} \| d\tau. \quad (9.13)$$

The term (9.10) tends to zero as  $t, s \rightarrow \infty$  by (9.4). Using (9.6) and (9.7) we see that the same holds for the terms in (9.11). The result holds trivially for the last term (9.13). We rewrite the term (9.12)

$$\int_s^t \tau^{-2} \| \nabla S(\tau) \cdot \nabla \hat{\phi} \| d\tau \leq c \left( \sup_{\tau \geq t_K} |\tau|^{-1/2} \| \nabla S(\tau) \|_{L^\infty(K)} \right) \| \nabla \hat{\phi} \| \int_s^t \tau^{-3/2} d\tau,$$

such that this term tends to zero as  $t, s \rightarrow \infty$  by (9.5).

We conclude that the limit

$$\varphi_+ = \lim_{t \rightarrow \infty} U(0, t) M(t) D(t) \exp(iS(t)) \hat{\phi}$$

exists. We have

$$\begin{aligned} M(t) D(t) \exp(iS(t)) \hat{\phi} &= M(t) D(t) F M(t) M(-t) F^{-1} \exp(iS(t)) \hat{\phi} \\ &= U_0(t) M(-t) \exp(iS(t, -i\nabla)) \varphi. \end{aligned}$$

and thus

$$\begin{aligned} & \| U_D(t) \varphi - M(t) D(t) \exp(iS(t)) \hat{\phi} \| \\ &= \| (1 - M(-t)) \exp(iS(t, -i\nabla)) \varphi \| \\ &\leq t^{-1/2} \| \nabla (\exp(iS(t)) \hat{\phi}) \| \\ &\leq t^{-1/2} \| \nabla S(t) \|_{L^\infty(K)} \| \hat{\phi} \| + t^{-1/2} \| \nabla \hat{\phi} \|. \end{aligned}$$

In the last step we have used  $|1 - \exp(i\xi^2/2t)| \leq |t|^{-1/2} |\xi|$  in the momentum representation. Using (9.5) we see that both terms tend to zero as  $t \rightarrow \infty$ . Putting everything together we have

$$\begin{aligned} \| U(0, t) U_D(t) \varphi - \varphi_+ \| &\leq \| U(0, t) (U_D(t) \varphi - M(t) D(t) \exp(iS(t)) \hat{\phi}) \| \\ &\quad + \| U(0, t) M(t) D(t) \exp(iS(t)) \hat{\phi} - \varphi_+ \|, \end{aligned}$$

and both terms tend to zero as  $t \rightarrow \infty$  by the above computations. This concludes the proof of the theorem.

## 10. Modified wave operators for Stark Hamiltonians

In this section we finally resolve the discrepancy between classical and quantum scattering for one-dimensional Stark Hamiltonians. We continue to use the notation

$$H_0 = -\frac{1}{2} \frac{d^2}{dx^2} + x, \quad H = H_0 + \Phi(x).$$

**Theorem 10.1** *Let  $\Phi \in \mathcal{B}^2(\mathbf{R})$  be a real valued function. Assume  $\Phi$  satisfies the following conditions:*

$$\lim_{x \rightarrow -\infty} \Phi(x) = 0, \quad (10.1)$$

$$\int_1^\infty (\|\chi_{\{x|x < -r^2\}} \Phi\|_\infty + \|\chi_{\{x|x < -r^2\}} \Phi'\|_\infty) r^{-1} dr < \infty. \quad (10.2)$$

Let  $S_G(t) = -\int_0^t \Phi(-\frac{1}{2}\tau^2) d\tau$ . Then the wave operators

$$W_\pm^G = s\text{-}\lim_{t \rightarrow \pm\infty} e^{itH} e^{-itH_0 + iS_G(t)} \quad (10.3)$$

exist. Furthermore,  $W_\pm^G$  satisfy the intertwining relation.

*Proof.* The proof goes through several steps. We first note that by the results in § 7 we can look at the time-dependent potential problem

$$K_0 = -\frac{1}{2} \frac{d^2}{dx^2}, \quad K(t) = -\frac{1}{2} \frac{d^2}{dx^2} + V(t, x) \quad (10.4)$$

with

$$V(t, x) = \Phi(x - \frac{1}{2}t^2). \quad (10.5)$$

By the discussion in § 7 there is a unique unitary propagator  $U(t, s)$  associated with  $K(t)$ . We can now extend the discussion in that section to Dollard-type modified wave operators and conclude that the limits in (10.3) exist if and only if the limits

$$\lim_{t \rightarrow \pm\infty} U(0, t) \exp(-itK_0 + iS_G(t)) \quad (10.6)$$

exist. Thus we can use the results in § 9 on Dollard-type modified wave operators for time-dependent perturbations of the Laplacian. The wave operators in (10.6) correspond to the decomposition  $V = V_s + V_l$  with

$$V_l(t, x) = \Phi(-\frac{1}{2}t^2), \quad (10.7a)$$

$$V_s(t, x) = \Phi(x - \frac{1}{2}t^2) - \Phi(-\frac{1}{2}t^2). \quad (10.7b)$$

We change this decomposition to the one used in [26]

$$V_l(t, x) = \Phi(x - \tfrac{1}{2}t^2), \quad (10.8a)$$

$$V_s(t, x) = 0, \quad (10.8b)$$

and let

$$S(t, y) = - \int_0^t \Phi(\tau y - \tfrac{1}{2}\tau^2) d\tau \quad (10.9)$$

as in §9. We then look at the existence of

$$W_{\pm}^D = s\text{-}\lim_{t \rightarrow \pm \infty} U(0, t) \exp(-itK_0 + iS(t, p)). \quad (10.10)$$

From now on we consider only the  $+$ -case. The two problems (10.6) and (10.10) are equivalent, if we can prove the following lemma.

*Lemma 10.2* Let  $K \subset \mathbf{R}$  be compact. Then

$$\Psi(y) = \lim_{t \rightarrow \infty} (S(t, y) - S_G(t))$$

exists, uniformly in  $y \in K$ .

*Proof of Lemma 10.2.* Assume  $R > 1$  is chosen such that  $K \subset [-R+1, R-1]$ . Assume  $t > R$  and  $y \in K$ . Then

$$S_G(t) - S(t, y) = \left( \int_0^R + \int_R^t \right) (\Phi(\tau y - \tfrac{1}{2}\tau^2) - \Phi(-\tfrac{1}{2}\tau^2)) d\tau.$$

The integral from 0 to  $R$  is a function of  $y$ , independent of  $t$ . In the second term we use integration by parts.

$$\begin{aligned} \int_R^t (\Phi(\tau y - \tfrac{1}{2}\tau^2) - \Phi(-\tfrac{1}{2}\tau^2)) d\tau &= \int_R^t \int_0^1 \tau y \Phi'(\tau \theta y - \tfrac{1}{2}\tau^2) d\theta d\tau \\ &= \int_0^1 \left( \frac{ty}{\theta y - t} \Phi(t\theta y - \tfrac{1}{2}t^2) - \frac{Ry}{\theta y - R} \Phi(R\theta y - \tfrac{1}{2}R^2) \right) d\theta \\ &\quad - \int_0^1 \int_R^t \left( \frac{y}{\theta y - \tau} + \frac{\tau y}{(\theta y - \tau)^2} \right) \Phi(\tau \theta y - \tfrac{1}{2}\tau^2) d\tau d\theta. \end{aligned}$$

By (10.1)

$$\frac{ty}{\theta y - t} \Phi(t\theta y - \tfrac{1}{2}t^2) \rightarrow 0 \quad \text{as } t \rightarrow \infty,$$

uniformly for  $y \in K$ ,  $\theta \in [0, 1]$ , so the first term in the first integral vanishes. The second term is independent of  $t$ . Furthermore,

$$\left| \left( \frac{y}{\theta y - \tau} + \frac{\tau y}{(\theta y - \tau)^2} \right) \Phi(\tau \theta y - \tfrac{1}{2}\tau^2) \right| \leq c \frac{1}{\tau} |\Phi(\tau \theta y - \tfrac{1}{2}\tau^2)|$$

Thus to prove the theorem it suffices to prove existence of the limits (10.10). We obtain this result as an application of Theorem 9.1. Hence we must verify the conditions in that theorem. Since  $V_s \equiv 0$ , the condition (i) is trivially satisfied. We now show that the function  $S(t, y)$  defined by (10.9) satisfies the conditions in (ii). Fix  $K \subset \mathbf{R}$  and  $R > 1$  such that  $K \subset [-R + 1, R - 1]$ . Take  $t_K = 4R + 1$ . Then  $S'(t, y) \in L^\infty(K)$  and  $S''(t, y) \in L^\infty(K)$  for all  $t \geq t_K$ . We have

$$-S'(t, y) = \int_0^t \tau \Phi'(\tau y - \frac{1}{2}\tau^2) d\tau = \left( \int_0^R + \int_R^t \right) \tau \Phi'(\tau y - \frac{1}{2}\tau^2) d\tau.$$

Now for  $y \in K$  and  $t > t_K$

$$\left| \int_0^R \tau \Phi'(\tau y - \frac{1}{2}\tau^2) d\tau \right| \leq R^2 \sup_{|x| \leq 2R^2} |\Phi'(x)|$$

and

$$\begin{aligned} \left| \int_R^t \tau \Phi'(\tau y - \frac{1}{2}\tau^2) d\tau \right| &= \left| -t(t-y)^{-1} \Phi(ty - \frac{1}{2}t^2) \right. \\ &\quad \left. + R(R-y)^{-1} \Phi(Ry - \frac{1}{2}R^2) \right. \\ &\quad \left. - \int_R^t y(y-\tau)^{-2} \Phi(\tau y - \frac{1}{2}\tau^2) d\tau \right| \\ &\leq (2+R) \|\Phi\|_\infty + R \|\Phi\|_\infty \int_R^\infty (\tau - R + 1)^{-2} d\tau. \end{aligned}$$

These computations imply

$$\|S'(t, \cdot)\|_{L^\infty(K)} \leq C, \quad t \geq t_K,$$

such that conditions (9.5) and (9.6) are satisfied. Similarly, we have

$$\begin{aligned} -S''(t, y) &= \int_R^t \tau^2 \Phi''(\tau y - \frac{1}{2}\tau^2) d\tau \\ &= t^2(y-t)^{-1} \Phi'(ty - \frac{1}{2}t^2) \\ &\quad - R^2(y-R)^{-1} \Phi'(Ry - \frac{1}{2}R^2) \\ &\quad + \{(y-t)^{-1} - y^2(y-t)^{-3}\} \Phi(ty - \frac{1}{2}t^2) \\ &\quad - \{(y-R)^{-1} - y^2(y-R)^{-3}\} \Phi(Ry - \frac{1}{2}R^2) \\ &\quad - \int_R^t ((y-\tau)^{-2} - 3y^2(y-\tau)^{-4}) \Phi(\tau y - \frac{1}{2}\tau^2) d\tau. \end{aligned}$$

We estimate

$$\left| \int_R^t \tau^2 \Phi''(\tau y - \frac{1}{2}\tau^2) d\tau \right| \leq 2t |\Phi'(ty - \frac{1}{2}t^2)| + c \sup_{|y| \leq 2R^2} |\Phi'(y)| + c \|\Phi\|_\infty.$$

and also

$$\left| \int_0^R \tau^2 \Phi''(\tau y - \frac{1}{2} \tau^2) d\tau \right| \leq R^3 \sup_{|x| \leq 2R^2} |\Phi''(x)|.$$

Combining these estimates we get

$$t^{-2} \|S''(t, \cdot)\|_{L^\infty(K)} \leq Ct^{-1} \|\chi_{\{|x| \leq -\frac{1}{4}t^2\}} \Phi'(x)\|_\infty + ct^{-2}$$

which is integrable by (10.2). Thus (9.7) holds.

Thus by Theorem 9.1 the limit (10.10) exists. The intertwining relation holds, since we have  $S_G(t+s) - S_G(t) \rightarrow 0$  as  $t \rightarrow \infty$  for each  $s \in \mathbf{R}$ . This concludes the proof of Theorem 10.1.

Theorem 10.1 should be compared with the classical mechanics results in §6.3. Now we have the same type of conditions in both cases, essentially  $\Phi(x) = O((\log|x|)^{-\beta})$  and  $\Phi'(x) = O((\log|x|)^{-\beta})$ ,  $\beta > 1$ , as  $x \rightarrow -\infty$ . The discrepancy is taken care of by the pure phase factor  $\exp(-i \int_0^t \Phi(-\frac{1}{2}\tau^2) d\tau)$  in the limits defining the quantum wave operators. A pure phase factor is not significant in the time evolution of wave packets under the free modified evolution.

We have not established asymptotic completeness for the wave operators under the conditions in Theorem 10.1, but we believe it holds. Assuming the stronger conditions  $\Phi \in C^\infty(\mathbf{R})$  and  $|\partial_x^k \Phi(x)| \leq c_k(1+|x|)^{-\varepsilon-k/2}$ ,  $k=0,1,2,\dots$ , completeness of  $W_\pm^D$  defined in (10.10) was proved in [26]. Combined with the above results completeness also holds for  $W_\pm^G$ . The result for  $W_\pm^G$  was also obtained in [44] under the same stronger conditions on the potential.

One may consider the same problem in dimensions  $d \geq 2$ . In [44] the problem is discussed under the conditions  $\Phi \in C^\infty(\mathbf{R}^d)$  and

$$|\partial_x^\alpha \Phi(x)| \leq c_\alpha(1+|x_1|)^{-\varepsilon-|\alpha|/2}$$

for all  $\alpha$ . Here we use the decomposition  $x = (x_1, x_\perp)$ , assuming that the electric field is in the  $x_1$ -direction. However, under this condition a nontrivial modification of the free evolution is needed in the variables  $x_\perp$ . On the other hand, in [12] existence and completeness of the  $W_\pm^G$  is proved in any dimension under the conditions  $\Phi \in \mathcal{B}^1(\mathbf{R}^d)$  and  $|\nabla \Phi(x)| \leq C(1+|x|)^{-1-\varepsilon}$ . One may ask whether these conditions are optimal.

In order to answer this question we need to generalize the results in Theorem 8.2 to include a modifier. We have the following result:

**Theorem 10.3.** Assume  $\Phi \in \mathcal{B}^2(\mathbf{R}^d)$  and  $F \in \mathbf{R}^d$ ,  $F \neq 0$ . Let  $V_t = \Phi(-\frac{1}{2}t^2 F)$  and  $V_s(t, x) = \Phi(x - \frac{1}{2}t^2 F) - \Phi(-\frac{1}{2}t^2 F)$ . Assume there exists  $\theta \in C(\mathbf{R}_+; \mathbf{R}_+)$ ,  $\int_0^\infty \theta(t) dt = +\infty$ , a closed set  $N$  of measure zero, and  $\mathcal{W} \in L_{\text{loc}}^2(\mathbf{R}^d \setminus N)$  such that for all compact  $K \subset \mathbf{R}^d \setminus N$

$$\lim_{t \rightarrow \infty} \|\chi_K(\theta(t)^{-1} D(t)^{-1} V_s(t) D(t) - \mathcal{W})\|_{\mathcal{B}(L^2(\mathbf{R}^d))} = 0.$$

Let  $\varphi \in L^2(\mathbf{R})$ . Assume

$$\lim_{t \rightarrow \infty} U(0, t) U_D(t) \varphi$$

*Proof.* We modify the argument in the proof of Theorem 8.2(i). Keeping track of the extra term  $\exp(iS(\tau))$ , which is independent of the  $x$ -variable, we find

$$(i\partial_\tau + \tfrac{1}{2}\Delta)M(\tau)D(\tau)\exp(iS(\tau))\hat{\psi} = M(\tau)D(\tau)\exp(iS(\tau))\left(\frac{1}{2\tau^2}\Delta + \Phi(-\tfrac{1}{2}\tau^2 F)\right)\hat{\psi}.$$

Note

$$D(\tau)^{-1}V(\tau)D(\tau) - \Phi(-\tfrac{1}{2}\tau^2 F) = D(\tau)^{-1}V_s(\tau)D(\tau).$$

Then we repeat the proof in Theorem 8.2, where we take  $\rho_K \equiv 0$  and

$$\sigma_K(t) = \|\chi_K(\theta(t)^{-1}D(t)^{-1}V_s(t)D(t) - \mathcal{W})\|_{\mathcal{B}(L^2(\mathbb{R}^d))}.$$

This completes the proof.

Using this result we can show that the result in [12] is optimal in dimensions  $d \geq 2$  in the following sense: Take  $d = 2$  and  $F = (2, 0)$ , and let

$$\Phi(x_1, x_2) = \frac{x_2}{(1 + x^2)^{1/2}(1 + \log(1 + x^2))}, \quad x = (x_1, x_2) \in \mathbb{R}^2.$$

Let  $N = \{0\}$ ,  $\theta(t) = (2t \log t)^{-1}$ ,  $t \geq 1$ , and  $\mathcal{W}(x_1, x_2) = x_2$ . Then it is easy to see that all the conditions in Theorem 10.3 are satisfied. We conclude that the Graf-modified wave operators fail to exist for this potential. Note that  $\nabla\Phi(x) = O(1/(|x|\log|x|))$  as  $|x| \rightarrow \infty$ . Hence the Graf result is optimal, and the modification in the  $x_\perp$ -variable for potentials with slower decay is needed.

### Acknowledgement

Many of the results in this paper were obtained in collaboration with E Mourre, P Perry, T Ozawa, and K Yajima. The results in this survey were presented at the Summer Workshop on Spectral and Inverse Spectral Theories, Kodaikanal, India, in August 1993. I want to thank Krishna Maddaly and Kalyan B Sinha for organizing a most interesting workshop. Preparation of these lecture notes was supported by the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine.

### References

- [1] Amrein W O, *Some questions in non-relativistic quantum scattering theory, scattering theory in mathematical physics* (eds) J A Lavita and J P Marchand (1974) (Dordrecht: D Reidel) pp. 97–140
- [2] Amrein W O, Nonrelativistic quantum dynamics, *Mathematical Physics Studies*, (1981) (Dordrecht: D Reidel) Vol. 2
- [3] Amrein W O, Jauch J M and Sinha K B, *Scattering theory in quantum mechanics* (Reading, Massachusetts) W A Benjamin (1977)
- [4] Avron J E and Herbst I W, Spectral and scattering theory for Schrödinger operators related to Stark effect, *Commun. Math. Phys.* **52** (1977) 239–254
- [5] Baumgärtel H and Wollenberg M, *Mathematical scattering theory, Operator theory: Advances and applications* (1983) (Basel: Birkhäuser Verlag) Vol. 9
- [6] Combes M, Spectral and scattering theory for a class of strongly oscillating potentials, *Commun. Math. Phys.* **73** (1980) 43–62
- [7] Cook J M, Convergence to the Møller wave matrix, *J. Math. Phys.* **36** (1957) 82–87

- [8] Cycon H L, Froese R G, Kirsch W and Simon B, *Schrödinger operators*, (1987) (Berlin: Springer Verlag)
- [9] Dollard J D, Asymptotic convergence and the Coulomb interaction, *J. Math. Phys.* **5** (1964) 729–738
- [10] Dunford N and Schwartz J T, *Linear operators. Part II: Spectral theory*, (1963) (New York: Interscience Publishers)
- [11] Enss V, Asymptotic completeness for quantum mechanical potential scattering, I. Short range potentials, *Commun. Math. Phys.* **61** (1978) 285–291
- [12] Graf G M, A remark on long-range Stark scattering, *Helv. Phys. Acta* **64** (1991) 1167–1174
- [13] Gustafson K, Candidates for  $\sigma_{ac}$  and  $H_{ac}$ , *Scattering Theory in Mathematical Physics* (eds) J A Lavita and J P Marchand (1974) (Dordrecht: D Reidel) pp. 157–168
- [14] Herbst I W, Unitary equivalence of Stark effect Hamiltonians, *Math. Z.* **155** (1977) 55–70
- [15] Herbst I W and Simon B, Dilation analyticity in constant electric fields. II. N-body problem. Borel summability, *Commun. Math. Phys.* **80** (1981) 181–216
- [16] Hörmander L, The existence of wave operators in scattering theory, *Math. Z.* **145** (1976) 69–91
- [17] Hörmander L, *The analysis of linear partial differential operators*, (1985) (Berlin: Springer Verlag) Vol. IV
- [18] Jensen A, Propagation estimates for Schrödinger-type operators, *Trans. Am. Math. Soc.* **291** (1985) 129–144
- [19] Jensen A, Asymptotic completeness for a new class of Stark effect Hamiltonians, *Commun. Math. Phys.* **107** (1986) 21–28
- [20] Jensen A, Commutator methods and asymptotic completeness for one-dimensional Stark effect Hamiltonians, Schrödinger operators. Aarhus 1985 (E. Balslev, ed.), Springer lecture notes in mathematics, vol. 1218, 1986, pp. 151–166
- [21] Jensen A, Scattering theory for Hamiltonians with Stark effect, *Ann. Inst. H. Poincaré, Phys. Théor.* **46** (1987) 383–395
- [22] Jensen A, Commutator methods and Schrödinger operators, *Rigorous results in quantum dynamics*, Liblice, Czechoslovakia, 10–15 June 1990 (eds) J Dittrich and P Exner (1991) (Singapore: World Scientific) pp. 3–15
- [23] Jensen A, Mourre E and Perry P, Commutator estimates and resolvent smoothness in quantum scattering theory, *Ann. Inst. H. Poincaré (N.S.)* **A41** (1984) 207–225
- [24] Jensen A and Ozawa T, Existence and non-existence results for wave operators for perturbations of the Laplacian, *Rev. Math. Phys.* **5** (1993), 601–629
- [25] Jensen A and Ozawa T, Classical and quantum scattering for Stark Hamiltonians with slowly decaying potentials, *Ann. Inst. H. Poincaré, Phys. Théor.* (1991), 229–243
- [26] Jensen A and Yajima K, On the long range scattering for Stark Hamiltonians, *J. reine angew. Math.* **420** (1991) 179–193
- [27] Kato T, Perturbation theory for linear operators, second ed., *Die Grundlehren der Mathematischen Wissenschaften*, (1976) (Berlin, Heidelberg, New York: Springer Verlag) Vol. 132
- [28] Kato T and Kuroda S T, Theory of simple scattering and eigenfunction expansions, *Functional Analysis and Related Fields* (1970) (Berlin, Heidelberg, New York: Springer-Verlag) pp. 99–131
- [29] Kuroda S T, On the existence and unitarity property of the scattering operator, *Nuovo Cimento* **12** (1959) 431–454
- [30] Kuroda S T, Scattering theory for differential operators, I. Operator theory, *J. Math. Soc. Japan* **25** (1973) 74–104
- [31] Littlewood J E, On the problem of  $n$  bodies, *Comm. Sem. Math. Lund* (1952), 143–155, tome suppl. dédié à M. Riesz
- [32] Mourre E, Link between the geometrical and the spectral transformation approaches in scattering theory, *Commun. Math. Phys.* **68** (1979) 91–94
- [33] Mourre E, Absence of singular continuous spectrum for certain self-adjoint operators, *Comm. Math. Phys.* **78** (1981) 391–408
- [34] Ozawa T, Non-existence of wave operators for Stark-effect Hamiltonians, *Math. Z.* **207** (1991) 335–339
- [35] Perry P A, Mellin transforms and scattering theory. I, short-range potentials, *Duke Math. J.* **47** (1980) 187–193
- [36] Perry P A, Scattering theory by the Enss method, *Math. Report* **1** (1983) 1–347
- [37] Reed M and Simon B, *Methods of modern mathematical physics. II: Fourier analysis. Selfadjointness*, (1975) (New York: Academic Press)
- [38] Reed M and Simon B, *Methods of modern mathematical physics. III: Scattering theory*, (1979) (New York: Academic Press)



- [39] Reed M and Simon B, *Methods of modern mathematical physics. I: Functional analysis*, revised edition ed. (1980) (New York: Academic Press)
- [40] Simon B, *Trace ideals and their applications* (1977) (Cambridge: Cambridge University Press)
- [41] Simon B, Phase space analysis of simple scattering systems: Extensions of some work of Enss, *Duke Math. J.* **46** (1979) 119–168
- [42] White D A W, Schrödinger operators with rapidly oscillating central potentials, *Trans. Am. Math. Soc.* **275** (1983) 641–677
- [43] White D A W, The Stark effect and long range scattering in two Hilbert spaces, *Indiana Univ. Math. J.* **39** (1990) 517–546
- [44] White D A W, Modified wave operators and Stark Hamiltonians, *Duke Math. J.* **68** (1992) 83–100
- [45] Yafaev D R, Mathematical scattering theory. General theory, *Translations of Mathematical Monographs* (1992) (Providence, Rhode Island: American Mathematical Society) vol. 105
- [46] Yajima K, Spectral and scattering theory for Schrödinger operators with Stark-effect, *J. Fac. Sci. Univ. Tokyo* **IA 26** (1979) 377–390
- [47] Yajima K, Spectral and scattering theory for Schrödinger operators with Stark-effect, II, *J. Fac. Sci. Univ. Tokyo* **IA 28** (1981) 1–15



# $L^p$ -Estimates for Schrödinger operators

SHU NAKAMURA

Department of Mathematical Sciences, University of Tokyo, 3-8-1, Komaba, Meguro, Tokyo, Japan 153

**Abstract.**  $L^p$ -estimates of the resolvent for a large class of Schrödinger operators are proved. Combining this with the almost analytic continuations, we obtain  $L^p$ -estimates for functions of Schrödinger operators.

**Keywords.** Schrödinger operators;  $L^p$ -estimates; functional calculus.

## 1. Introduction

The purpose of these lectures is to present some recent results on  $L^p$ -estimates for Schrödinger operators. In particular, we will discuss estimates for the resolvents and, more generally, functions of Schrödinger operators in  $L^p$  with  $1 \leq p \leq \infty$ . Most of the result in these lectures are outcome of the joint work of Arne Jensen and the author [JN1, JN2], and presented in a slightly generalized form.

We consider Schrödinger operators:  $H = -\Delta + V(x)$ , which is primarily defined as a self-adjoint operator on  $L^2(\mathbf{R}^d)$  with  $d \geq 1$ . We always suppose that the potential  $V(x)$  satisfies the following assumption:

*Assumption.*  $V(x)$  is a real-valued function on  $\mathbf{R}^d$  such that  $V(x) = V_+(x) - V_-(x)$ ,  $V_\pm(x) \geq 0$ ,  $V_+ \in L^1_{\text{loc}}(\mathbf{R}^d)$  and  $V_- \in K_d$ , where  $K_d$  is the Kato-class of functions defined by:  $V \in K_d$  if

(i) If  $d = 1$ ,

$$\sup_{x \in \mathbf{R}} \int_{|x-y| \leq 1} |V(y)| dy < \infty;$$

(ii) If  $d = 2$ ,

$$\lim_{\alpha \downarrow 0} \left[ \sup_{x \in \mathbf{R}^2} \int_{|x-y| \leq \alpha} \log(|x-y|^{-1}) |V(y)| dy \right] = 0;$$

(iii) If  $d \geq 3$ ,

$$\lim_{\alpha \downarrow 0} \left[ \sup_{x \in \mathbf{R}^d} \int_{|x-y| \leq \alpha} \frac{|V(y)|}{|x-y|^{d-2}} dy \right] = 0.$$

Then it is well-known that the Friedrichs extension with the form domain  $\mathcal{Q}(H) = H^1(\mathbf{R}^d) \cap D(|V|^{1/2})$  exists, and it is bounded from below, i.e.,  $H \geq C$  with some

$C \in \mathbf{R}$ . Hence we can define the heat semigroup generated by  $H$ , often called the *Schrödinger semigroup*, by  $e^{-tH}$  for  $t \geq 0$ .

Under our assumption, Aizenman and Simon [AS] showed that the Feynman-Kac formula holds for the Schrödinger semigroup:

$$(e^{-tH}\psi)(x) = \int_{\Omega} \exp\left(-\int_0^t V(\omega(s))ds\right) \psi(\omega(t)) d\mu_x(\omega),$$

where  $\mu_x$  is the Wiener measure (Brownian motion) starting at  $x$  at  $t=0$ , and  $\Omega$  is the space of all continuous paths on  $\mathbf{R}^d$ .

Based on the Feynman-Kac formula, Simon investigated  $L^p$ -properties of Schrödinger operators in a well-known monograph on Schrödinger semigroup [S]. For example, using the Khasmin'skii lemma, he showed that if  $1 \leq p \leq q \leq \infty$ , then

$$\|e^{-tH}\|_{B(L^p, L^q)} \leq Ce^{At} t^{-\gamma}, \quad t > 0, \quad (1.1)$$

with some  $A, C > 0$  and  $\gamma = (d/2)(1/p - 1/q)$ . In particular,

$$\|e^{-tH}\|_{B(L^p)} \leq Ce^{At}, \quad t > 0, \quad (1.2)$$

with some  $A \in \mathbf{R}$ . It is easy to see that  $\{e^{-tH}, t \geq 0\}$  is a strongly continuous semigroup in  $L^p(\mathbf{R}^d)$ , since  $L^2 \cap L^p$  is dense in  $L^p(\mathbf{R}^d)$ . We can thus define  $H$  on  $L^p(\mathbf{R}^d)$  as the generator of the semigroup in  $L^p(\mathbf{R}^d)$ . The above estimate then implies that the  $L^p$ -spectrum (the spectrum in  $L^p$ -space) is included in the half-space:  $\sigma_{L^p}(H) \subset \{z | \operatorname{Im} z \geq -A\}$  since

$$(H - z)^{-1} = \int_0^\infty e^{-tH} e^{zt} dt \in B(L^p), \quad \text{if } \operatorname{Im} z > -A, \quad (1.3)$$

where  $B(X)$  is the space of the bounded operators in  $X$ . In fact, we can take any  $A > -\inf \sigma(H)$  for any  $p \in [1, \infty]$ , where  $\sigma(H) \equiv \sigma_{L^2}(H)$ .

He suggested the invariance of the spectrum:  $\sigma_{L^p}(H) = \sigma(H)$  for any  $p$  as an open question, and it was later proved by Hempel and Voigt [HV]. They used the Combes-Thomas boost technic [CT] to show the exponential decay of the integral kernel of  $(H - z)^{-1}$ . Namely, if  $z \notin \sigma(H)$ , then

$$|(H - z)^{-1}(x, y)| \leq C_z \exp(-\gamma_z |x - y|), \quad \text{for } |x - y| > 1,$$

with some  $\gamma_z > 0$ . Then this implies  $(H - z)^{-1} \in B(L^p(\mathbf{R}^d))$ , and hence  $z \notin \sigma_{L^p}(H)$  for any  $p$ . The converse:  $\sigma_{L^p}(H) \supset \sigma_{L^2}(H)$  is easy, and they imply  $\sigma_{L^p}(H) = \sigma(H)$ .

In these lectures, we prove improved (probably optimal) estimates for the resolvent in  $L^p(\mathbf{R}^d)$  using the function space  $\ell^p(L^2)$ , sometimes called the *amalgam* of  $\ell^p$  and  $L^2$ . Then we use the machinery of the almost analytic continuation to obtain  $L^p$ -estimates for functions of the Schrödinger operator  $f(H)$ , which is defined in  $L^2(\mathbf{R}^d)$  by the functional calculus at first, and then extended to an operator in  $L^p(\mathbf{R}^d)$ . Thus, in a sense, we can construct a theory of functional calculus in  $L^p$ -space. This idea is employed and generalized by Davies [D], where he constructed an abstract theory of functional calculus in Banach space.

# $L^p$ -estimates for the resolvent

goal of this section is the following:

**orem 1.** *Let  $1 \leq p \leq \infty$ . Then for any  $R > 0$ ,*

$$\|(H - z)^{-1}\|_{B(L^p)} \leq C \operatorname{dist}(z, \sigma(H))^{-\beta-1}, \quad \text{if } |z| \leq R, \quad (2.1)$$

here  $\beta = d|1/2 - 1/p|$ .

In order to prove Theorem 1, we introduce the space  $\ell^p(L^q)$ :

$$\ell^p(L^q) = \left\{ \varphi \in L^q_{\text{loc}}(\mathbf{R}^d) \left| \sum_{n \in \mathbf{Z}^d} \|\varphi\|_{L^q(C(n))}^p < \infty \right. \right\}$$

$1 \leq p < \infty$ , where  $C(n) = \{x \in \mathbf{R}^d | |x_i - n_i| \leq 1/2, i = 1, \dots, d\}$  is the unit cube at  $n \in \mathbf{Z}^d$ .

The modification for the case  $p = \infty$  is obvious.  $\ell^p(L^q)$  is a Banach space with the norm:

$$\|\varphi\|_{\ell^p(L^q)} = \left( \sum_{n \in \mathbf{Z}^d} \|\varphi\|_{L^q(C(n))}^p \right)^{1/p},$$

$\varphi \in \ell^p(L^q)$ .

We first note an inequality of Young's type:

**mma 2.** *Let  $1 \leq p, q, r, s, t, u \leq \infty$  such that  $1/p + 1/q - 1 = 1/r$  and  $1/s + 1/t - 1 = 1/u$ . If  $f \in \ell^p(L^s)$  and  $g \in \ell^q(L^t)$  then  $f * g \in \ell^r(L^u)$  and*

$$\|f * g\|_{\ell^r(L^u)} \leq 3^d \|f\|_{\ell^p(L^s)} \|g\|_{\ell^q(L^t)}. \quad (2.2)$$

The proof just uses Young's inequalities in  $L^p$ -space and  $\ell^p$ -space sequentially to estimate the left hand side. We omit the details. Combining this with a kernel estimate for the Schrödinger semigroup, we can show the following mapping property of  $e^{-tH}$  between  $L^p$ -spaces and  $\ell^p(L^q)$ -spaces.

**mma 3.** *Let  $1 \leq p \leq q \leq \infty$ . Then  $e^{-tH}$  is bounded from  $L^p(\mathbf{R}^d)$  to  $\ell^p(L^q)$  for each  $t > 0$ , and*

$$\|e^{-tH}\|_{B(L^p, \ell^p(L^q))} \leq C e^{Lt} (1 + t^{-d(1/p - 1/q)/2}), \quad \text{for } t > 0 \quad (2.3)$$

for some  $C, L > 0$ .

*Sketch of Proof.* We first note the kernel estimate for  $e^{-tH}$ : for any  $\varepsilon > 0$ ,

$$e^{-tH}(x, y) \leq C e^{At} t^{-d/2} \exp\left(-\frac{|x - y|^2}{(4 + \varepsilon)t}\right) \equiv G(t, x - y)$$

for  $x, y \in \mathbf{R}^d$  and  $t > 0$  (see [S], Proposition B.6.7). Then we compute  $\|G(t, \cdot)\|_{\ell^1(L^p)}$  using this. By simple computations, we obtain

$$\sum_{n \neq 0} \|G(t, \cdot)\|_{L^p(C(n))} \leq C e^{At}.$$

Hence we have

$$\|G(t, \cdot)\|_{\ell^1(L^p)} \leq C e^{At} (1 + t^{-d(1-1/p)/2}).$$

Now we use Lemma 2 to obtain

$$\begin{aligned} \|e^{-tH} \varphi\|_{\ell^p(L^q)} &\leq \|G(t, \cdot) * |\varphi|\|_{\ell^p(L^q)} \leq 3^d \|G(t, \cdot)\|_{\ell^1(L^r)} \|\varphi\|_{\ell^p(L^p)} \\ &\leq C e^{Lt} (t^{-d(1-1/r)/2} + 1) \|\varphi\|_{L^p} \\ &= C e^{Lt} (t^{-d(1/p-1/q)/2} + 1) \|\varphi\|_{L^p}, \end{aligned}$$

where

$$1/r + 1/p - 1 = 1/q.$$

The following lemma is the key to reduce our  $L^p$ -problem to an  $\ell^p(L^q)$ -problem.

*Lemma 4.* Let  $1 \leq p \leq q \leq \infty$  and let  $\beta > d(1/p - 1/q)/2$ . Then for sufficiently large  $(H + M)^{-\beta}$  is bounded from  $L^p(\mathbf{R}^d)$  to  $\ell^p(L^q)$ .

*Proof.* We use the well-known formula: for  $\beta > 0$ ,

$$(H + M)^{-\beta} = \frac{1}{\Gamma(\beta)} \int_0^\infty e^{-tH} e^{-Mt} t^{\beta-1} dt.$$

We take  $M > L$  in Lemma 3 and use Lemma 3 to obtain

$$\|(H + M)^{-\beta} \varphi\|_{\ell^p(L^q)} \leq \frac{C}{\Gamma(\beta)} \int_0^\infty e^{-(M-L)t} (1 + t^{-d(1/p-1/q)/2}) t^\beta \frac{dt}{t} \|\varphi\|_{L^p} < \infty.$$

We now turn to the problem of boundedness in  $\ell^1(L^2)$ -space.

**DEFINITION.**

$A \in \mathcal{A}_\beta$  if  $A \in B(L^2(\mathbf{R}^d))$  and

$$\sup_{n \in \mathbf{Z}^d} \|\langle \cdot - n \rangle^\beta A \chi_{C(n)}\|_{B(L^2)} < \infty.$$

where  $\chi_\Omega$  denotes the indicator function of  $\Omega$ . The norm of  $A$  in  $\mathcal{A}_\beta$  is defined by

$$\|A\|_\beta \equiv \|A\|_{B(L^2)} + \sup_{n \in \mathbf{Z}^d} \|\langle \cdot - n \rangle^\beta A \chi_{C(n)}\|_{B(L^2)}.$$

This class of operators is useful in studying the boundedness of operators in  $\ell^1(L^2)$ . In fact, if  $A \in \mathcal{A}_\beta$  with  $\beta > d/2$ , then  $A$  is bounded in  $\ell^1(L^2)$ .

**Theorem 5.** If  $A \in \mathcal{A}_\beta$  with  $\beta > d/2$ , then  $A \in B(\ell^1(L^2))$  and

$$\|A\|_{B(\ell^1(L^2))} \leq C \|A\|_\beta^{d/2\beta} \|A\|^{1-d/2\beta}, \quad (2.4)$$

where  $C$  depends only on  $d$  and  $\beta$ .

*Proof.* We write  $\chi_n = \chi_{C(n)}(x)$ . If  $A \in \mathcal{A}_\beta$ , then by the definition,

$$\left( \sum_{m \in \mathbb{Z}^d} \langle m - n \rangle^{2\beta} \|\chi_m A \chi_n \varphi\|^2 \right)^{1/2} \leq C \|\langle \cdot - n \rangle^\beta A \chi_n \varphi\| \leq C \|A\|_\beta \|\chi_n \varphi\|$$

We remark that if we use the Schwarz inequality here, we obtain

$$\begin{aligned} \sum_m \|\chi_m A \chi_n \varphi\| &\leq \left( \sum_m \langle m - n \rangle^{-2\beta} \right)^{1/2} \left( \sum_m \langle m - n \rangle^{2\beta} \|\chi_m A \chi_n \varphi\|^2 \right)^{1/2} \\ &\leq C \|A\|_\beta \|\chi_n \varphi\|. \end{aligned}$$

By the definition of  $\ell^1(L^2)$ , this implies  $\|A \chi_n \varphi\|_{\ell^1(L^2)} \leq C \|A\|_\beta \|\chi_n \varphi\|$ . Then by the triangle inequality, we have

$$\|A \varphi\|_{\ell^1(L^2)} \leq \sum_n \|A \chi_n \varphi\|_{\ell^1(L^2)} \leq C \|A\|_\beta \sum_n \|\chi_n \varphi\| = C \|A\|_\beta \|\varphi\|_{\ell^1(L^2)}.$$

But this is not sufficient for our purpose, and we use the following trick.

Let  $\omega > 0$ . Then

$$\begin{aligned} &\sum_m \|\chi_m A \chi_n \varphi\| \\ &= \sum_{|n-m| > \omega} |m-n|^{-\beta} |m-n|^\beta \|\chi_m A \chi_n \varphi\| + \sum_{|n-m| \leq \omega} \|\chi_m A \chi_n \varphi\| \\ &\leq \left( \sum_{|m-n| > \omega} |m-n|^{-2\beta} \right)^{1/2} \left( \sum_m |m-n|^{2\beta} \|\chi_m A \chi_n \varphi\|^2 \right)^{1/2} \\ &\quad + \left( \sum_{|n-m| \leq \omega} 1 \right)^{1/2} \left( \sum_m \|\chi_m A \chi_n \varphi\|^2 \right)^{1/2} \\ &\leq C \left( \int_{|x| > \omega} |x|^{-2\beta} dx \right)^{1/2} \|A\|_\beta \|\chi_n \varphi\| + C \omega^{d/2} \|A\| \|\chi_n \varphi\| \\ &\leq C (\omega^{-(\beta-d/2)} \|A\|_\beta + \omega^{d/2} \|A\|) \|\chi_n \varphi\|. \end{aligned}$$

by the Schwarz inequality. Now we set  $\omega = (\|A\|_\beta / \|A\|)^{1/\beta}$  so that the two terms in the last expression are the same order in  $\|A\|_\beta$  and  $\|A\|$ . Then we have

$$\sum_m \|\chi_m A \chi_n \varphi\| \leq 2C \|A\|_\beta^{d/2\beta} \|A\|^{1-d/2\beta} \|\chi_n \varphi\|.$$

Summing up this in  $n$ , we obtain

$$\|A \varphi\|_{\ell^1(L^2)} \leq C \|A\|_\beta^{d/2\beta} \|A\|^{1-d/2\beta} \|\varphi\|_{\ell^1(L^2)}.$$

■

Then we check if  $(H - z)^{-1} \in \mathcal{A}_\beta$ . It turns out that it is fairly simple.

**Lemma 6.** Let  $R > 0$  and let  $l > 0$ . Then there exists  $C > 0$  such that

$$\sup_{n \in \mathbb{Z}^d} \|\langle \cdot - n \rangle^l (H - z)^{-1} \langle \cdot - n \rangle^{-l}\| \leq C \operatorname{dist}(z, \sigma_{L^2}(H))^{-l-1} \quad (2.5)$$

if  $z \notin \sigma_{L^2}(H)$  and  $|z| \leq R$ .

*Proof.* For simplicity, we consider the case  $l = 1$  only. We compute the commutator:

$$[x, (H - z)^{-1}] = (H - z)^{-1} [H, x] (H - z)^{-1} = 2(H - z)^{-1} \left( \frac{\partial}{\partial x} \right) (H - z)^{-1}$$

as a quadratic form on  $C_0^\infty(\mathbb{R}^d)$ . Under our assumption, the differential operator  $(\partial/\partial x)$  is  $H$ -bounded, and hence

$$\begin{aligned} \left\| \left( \frac{\partial}{\partial x} \right) (H - z)^{-1} \varphi \right\| &\leq C(\|H(H - z)^{-1} \varphi\| + \|(H - z)^{-1} \varphi\|) \\ &\leq C(1 + (1 + |z|)\|(H - z)^{-1}\|)\|\varphi\| \\ &\leq C \operatorname{dist}(z, \sigma(H))^{-1} \|\varphi\| \quad \text{if } |z| \leq R. \end{aligned}$$

Hence we obtain

$$\|[(x - n), (H - z)^{-1}]\| \leq C \operatorname{dist}(z, \sigma(H))^{-2} \quad \text{if } |z| \leq R.$$

This implies

$$\begin{aligned} \|(x - n)(H - z)^{-1} \langle x - n \rangle^{-1}\| &\leq \|(H - z)^{-1}\| + \|[(x - n), (H - z)^{-1}]\| \\ &\leq C \operatorname{dist}(z, \sigma(H))^{-2} \quad \text{if } |z| \leq R. \end{aligned}$$

Repeating this procedure, we can prove the estimate for any  $l$ . ■

**Lemma 7.** For any  $\beta \geq 0$  and  $z \notin \sigma_{L^2}(H)$ ,  $(H - z)^{-1} \in \mathcal{A}_\beta$ . Moreover, for any  $R > 0$  and  $\beta > 0$ ,

$$\|(H - z)^{-1}\|_\beta \leq C \operatorname{dist}(z, \sigma_{L^2}(H))^{-\beta-1} \quad \text{for } |z| \leq R. \quad (2.6)$$

If  $\beta$  is an integer, Lemma 7 is a direct consequence of Lemma 6, since

$$\|\langle \cdot - n \rangle^l (H - z)^{-1} \chi_{C(n)}\| \leq C \|\langle \cdot - n \rangle^l (H - z)^{-1} \langle \cdot - n \rangle^{-l}\|.$$

For general  $\beta$ , the assertion follows by complex interpolation.

*Proof of Theorem 1.* We let  $m > d/4$  be an integer. By the resolvent equation, it is easy to see

$$(H - z)^{-1} = \sum_{k=0}^m (z + M)^{k-1} (H + M)^{-k} + (z + M)^m (H - z)^{-1} (H + M)^{-m}$$



hand,

$$\begin{aligned} & \| (H - z)^{-1} (H + M)^{-m} \|_{B(L^1)} \\ & \leq \| (H - z)^{-1} \|_{B(\ell^1(L^2), L^1)} \| (H + M)^{-m} \|_{B(L^1, \ell^1(L^2))} \\ & \leq C \| (H - z)^{-1} \|_{B(\ell^1(L^2))} \leq C \| (H - z)^{-1} \|_{\beta}^{d/2\beta} \| (H - z)^{-1} \|^{1-d/2\beta} \end{aligned}$$

if  $\beta > d/2$  by Theorem 5. Here we have used the fact that  $\ell^1(L^2)$  is continuously embedded in  $L^1(\mathbf{R}^d)$ . Then, by Lemma 7, we learn

$$\begin{aligned} & \| (H - z)^{-1} (H + M)^{-m} \|_{B(L^1)} \\ & \leq C [\text{dist}(z, \sigma_{L^2}(H))^{-1-\beta}]^{d/2\beta} [\text{dist}(z, \sigma_{L^2}(H))^{-1}]^{1-d/2\beta} \\ & = C \text{dist}(z, \sigma_{L^2}(H))^{-1-d/2} \quad \text{if } |z| \leq R. \end{aligned}$$

This completes the proof of Theorem 1 for  $p = 1$ . On the other hand, it is clear that  $\| (H - z)^{-1} \| \leq \text{dist}(z, \sigma(H))^{-1}$ . Combining them with the complex interpolation, we obtain the assertion for  $1 \leq p \leq 2$ . The other cases follow by the duality argument. ■

In the next section, we need global estimates for the resolvent in  $L^p(\mathbf{R}^d)$ , whereas Theorem 1 is local. Theorem 1 implies, in particular,

$$\| (H - z)^{-1} \|_{B(L^p)} \leq C |\text{Im } z|^{-1-\beta}, \quad \text{if } |z| \leq R$$

with  $\beta = d|1/2 - 1/p|$ . We can prove that the above estimate holds globally. Namely,

**Theorem 8.** *Let  $1 \leq p \leq \infty$  and let  $\beta = d|1/2 - 1/p|$ . Then*

$$\| (H - z)^{-1} \|_{B(L^p)} \leq C \frac{\langle z \rangle^\beta}{|\text{Im } z|^{\beta+1}} \quad \text{for } z \in \mathbf{R} \setminus \mathbf{C}. \quad (2.7)$$

*Sketch of Proof.* We use the scaling:

$$U_p(\theta)\varphi(x) = \theta^{d/p}\varphi(\theta x), \quad \text{for } \theta \in (0, 1] \quad \text{and} \quad \varphi \in L^p(\mathbf{R}^d).$$

$U_p(\theta)$  is an isometry in  $L^p(\mathbf{R}^d)$ . Moreover, it is easy to see

$$U_p(\sqrt{\theta}) H U_p(\sqrt{\theta})^{-1} = \theta^{-1}(-\Delta) + V(\sqrt{\theta}x).$$

This implies

$$\theta H = U_p(\sqrt{\theta})^{-1} H_\theta U_p(\sqrt{\theta}), \quad \text{where } H_\theta = -\Delta + \theta V(\sqrt{\theta}x).$$

We can show that all the estimates for  $H$  so far holds for  $H_\theta$  uniformly in  $\theta \in (0, 1]$ . Thus,

$$\| (H_\theta - z)^{-1} \|_{B(L^p)} \leq C |\text{Im } z|^{-\beta-1} \quad \text{if } |z| \leq 1 \quad \text{and} \quad \theta \in (0, 1].$$

For  $|z| > 1$ , we set  $\hat{z} = z/|z|$  and  $\theta = 1/|z|$ . Then

$$\begin{aligned} \| (H - z)^{-1} \|_{B(L^p)} &= |z|^{-1} \| (\theta H - \hat{z})^{-1} \|_{B(L^p)} = |z|^{-1} \| (H_\theta - \hat{z})^{-1} \|_{B(L^p)} \\ &\leq C |z|^{-1} |\text{Im } \hat{z}|^{-\beta-1} = C \frac{|z|^\beta}{|\text{Im } z|^{\beta+1}}. \end{aligned}$$

■

In this section, we explain the construction of *almost analytic continuations* for functions on  $\mathbf{R}$ , and combining this with the resolvent estimates in the last section, we prove several  $L^p$ -estimates for functions of Schrödinger operators  $f(H)$ .

### PROPOSITION 9.

Let  $k \geq 1$  and let  $\alpha \in \mathbf{R}$ . Suppose  $f \in C^{k+1}(\mathbf{R})$  such that

$$\left| \left( \frac{d}{d\lambda} \right)^j f(\lambda) \right| \leq C \langle \lambda \rangle^{\alpha-j} \quad \text{for } j = 0, 1, \dots, k+1 \quad \text{and } \lambda \in \mathbf{R}. \quad (3.1)$$

Then there exists  $\tilde{f} \in C^1(\mathbf{C})$  such that  $\tilde{f}(\lambda) = f(\lambda)$  for  $\lambda \in \mathbf{R}$  and

$$|\partial_{\bar{z}} \tilde{f}(z)| \leq C |\operatorname{Im} z|^j \langle z \rangle^{\alpha-1-j} \quad \text{for } z \in \mathbf{C} \quad \text{and } j = 0, 1, \dots, k, \quad (3.2)$$

where

$$(\partial_{\bar{z}} f)(x + iy) = \frac{1}{2} (\partial_x f + i \partial_y f)(x + iy).$$

*Proof.* We set

$$f_1(x + iy) = \sum_{j=0}^k \frac{(iy)^j}{j!} f^{(j)}(x) \quad \text{for } x, y \in \mathbf{R}.$$

Then it is easy to see

$$\begin{aligned} \partial_{\bar{z}} f_1(x + iy) &= \frac{1}{2} \sum_{j=0}^k \left[ \frac{(iy)^j}{j!} f^{(j+1)}(x) - \frac{j}{j!} (iy)^{j-1} f^{(j)}(x) \right] \\ &= \frac{1}{2} \frac{(iy)^k}{k!} f^{(k+1)}(x). \end{aligned}$$

We let  $\psi \in C^\infty(\mathbf{C})$  such that

$$\psi(z) = \begin{cases} 1 & \text{if } |\operatorname{Im} z| \leq 1 + |\operatorname{Re} z|, \\ 0 & \text{if } |\operatorname{Im} z| \geq 2 + 2|\operatorname{Re} z|, \end{cases}$$

and  $\psi(z) = \tilde{\psi}(z/|z|)$  for  $|z| \gg 1$  with some  $\tilde{\psi}$ . We now set  $\tilde{f}(z) = \psi(z) f_1(z)$ . It remains only to check the decay properties, and it is straightforward. ■

If  $f \in C^\infty(\mathbf{R})$ , we can expect  $\tilde{f} \in C^\infty(\mathbf{C})$ . Usually this is proved using the asymptotic sum (or Borel sum), but we can also use more direct construction.

### PROPOSITION 10.

Let  $f \in C^\infty(\mathbf{R})$  such that

$$\left| \left( \frac{d}{d\lambda} \right)^j f(\lambda) \right| \leq C_j \langle \lambda \rangle^{\alpha-j} \quad \text{for } \lambda \in \mathbf{R}$$

some  $\alpha \in \mathbf{R}$  and all  $j \geq 0$ . Then there exists  $\tilde{f} \in C^\infty(\mathbf{C})$  such that

$$|\partial_{\bar{z}} \tilde{f}(z)| \leq C_N |\operatorname{Im} z|^N \langle z \rangle^{\alpha-1-N} \quad \text{for } z \in \mathbf{C} \quad \text{and all } N \geq 0. \quad (3.3)$$

*Sketch of Proof.* We first consider  $f \in C_0^\infty(-2, 2)$ . Let  $\chi \in C_0^\infty(\mathbf{R})$  such that

$$\chi(\lambda) = \begin{cases} 1 & \text{if } |\lambda| \leq 1, \\ 0 & \text{if } |\lambda| \geq 2, \end{cases}$$

and  $0 \leq \chi(\lambda) \leq 1$ , and let  $\rho(\lambda) = \int_0^\lambda \chi(\mu) d\mu$ . We set

$$(Tf)(x + iy) = (2\pi)^{-1/2} \chi(x/2) \chi(y) \int_0^\infty e^{-\rho(y\xi)} e^{ix\xi} \hat{f}(\xi) d\xi,$$

where  $\hat{f}$  is the Fourier transform of  $f$ , i.e.,  $\hat{f}(\xi) = (2\pi)^{-1/2} \int_{-\infty}^\infty e^{-ix\xi} f(x) dx$ . Since  $\rho(y\xi) \in C^\infty(\mathbf{R})$  is bounded, it is easy to see  $Tf \in C_0^\infty(\mathbf{C})$  and,

$$\operatorname{supp} Tf \subset \{z \mid |\operatorname{Re} z| \leq 2, |\operatorname{Im} z| \leq 1\}.$$

Moreover, since  $\rho(\lambda) = \lambda$  for  $|\lambda| \leq 1$ ,

$$\partial_{\bar{z}}(e^{-\rho(y\xi)} e^{ix\xi}) = 0 \quad \text{if } |y| \cdot |\xi| \leq 1.$$

Hence

$$\begin{aligned} \left| \partial_{\bar{z}} \int_{-\infty}^\infty e^{-\rho(y\xi)} e^{ix\xi} \hat{f}(\xi) d\xi \right| &\leq \frac{1}{2} \int |i\xi(1 - \rho'(y\xi))| |\hat{f}(\xi)| d\xi \\ &\leq C \int |\xi|^{N+1} |y|^N |\hat{f}(\xi)| d\xi \leq C |y|^N. \end{aligned}$$

We can also estimate the other terms to obtain

$$|\partial_{\bar{z}}(Tf)(z)| \leq C_N |z|^N \quad \text{for any } N.$$

For general  $f$ , we use the standard method of partition of unity. We let  $\varphi \in C_0^\infty(1/2, 2)$  that

$$\sum_{k=-\infty}^\infty \varphi(2^k \lambda) = 1 \quad \text{for any } \lambda > 0.$$

Then we set  $\varphi_{\pm k}(\lambda) = \varphi(\pm 2^{-k} \lambda)$  for  $k = 1, 2, \dots$ , and  $\varphi_0(\lambda) = 1 - \sum_{j \neq 0} \varphi_j(\lambda)$ . We compose  $f$  by these cut-off functions:

$$f(\lambda) = \sum_j f(\lambda) \varphi_j(\lambda) = \sum_j f_j(2^{-|j|} \lambda),$$

where  $f_j(\lambda) = \varphi(\operatorname{sgn}(j) \lambda) f(2^{|j|} \lambda)$  for  $k > 0$  and  $f_0(\lambda) = \varphi_0(\lambda) f(\lambda)$ . Clearly  $f_j \in C_0^\infty(-2, 2)$ , and we can apply the above construction of  $\tilde{f}$  to each  $f_j$  to obtain  $\tilde{f}(z) = \sum_j (Tf_j)(2^{-|j|} z)$ . It is easy to see  $\tilde{f}(z)$  satisfies the claim.  $\blacksquare$

## DEFINITION

A Propositions 9 and 10 is called an *almost analytic continuation* of  $f$ .

We use the following useful formula of  $f(H)$  in terms of the almost analytic continuation of  $f$ , probably due to Helffer and Sjöstrand [HS].

# PROPOSITION 11

Let  $f \in C^2(\mathbf{R})$  and satisfies

$$\left| \left( \frac{d}{d\lambda} \right)^j f(\lambda) \right| \leq C \langle \lambda \rangle^{-\varepsilon-j} \quad \text{for } \lambda \in \mathbf{R} \quad \text{and } j = 0, 1, 2,$$

with some  $\varepsilon > 0$ . Let  $A$  be a self-adjoint operator in  $\mathcal{H}$ . Then

$$f(A) = \frac{1}{2\pi i} \int_{\mathbf{C}} (\partial_{\bar{z}} \tilde{f}(z)) (A - z)^{-1} dz d\bar{z}, \quad (3.4)$$

where  $\tilde{f}$  is an almost analytic continuation of  $f$ .

Actually the proof is elementary, but we prove it for the completeness.

**Lemma 12. (Generalized Cauchy's theorem)** Let  $\Omega \subset \mathbf{C}$  be a bounded domain and let  $g \in C^2(\bar{\Omega})$ . Then

$$\int_{\partial\Omega} g(z) dz = - \int_{\Omega} \partial_{\bar{z}} g dz d\bar{z}. \quad (3.5)$$

*Proof.* We use the Stokes' formula for  $\Omega \subset \mathbf{C} \cong \mathbf{R}^2$ ,  $z = x + iy$ ,  $x, y \in \mathbf{R}$ . Then  $dz = dx + idy$  and hence

$$\begin{aligned} d(g(z) dz) &= \left( \frac{\partial g}{\partial x} dx + \frac{\partial g}{\partial y} dy \right) \wedge (dx + idy) = \left( i \frac{\partial g}{\partial x} - \frac{\partial g}{\partial y} \right) dx \wedge dy \\ &= i \left( \frac{\partial g}{\partial x} + i \frac{\partial g}{\partial y} \right) dx \wedge dy = 2i \partial_{\bar{z}} g dx \wedge dy. \end{aligned}$$

On the other hand,

$$dz d\bar{z} = (dx + idy) \wedge (dx - idy) = -2i dx \wedge dy,$$

Hence we obtain  $d(g dz) = -\partial_{\bar{z}} g dz d\bar{z}$ . Using this and Stokes' formula, we have

$$\int_{\partial\Omega} g dz = \int_{\Omega} d(g dz) = - \int_{\Omega} \partial_{\bar{z}} g dz d\bar{z}. \quad \blacksquare$$

Now using the same argument to prove Cauchy's formula from Cauchy's theorem, we obtain the following formula by Lemma 12:

**Lemma 13.** Let  $\Omega \subset \mathbf{C}$  be a bounded domain and let  $\lambda \in \Omega$ . Let  $g \in C^2(\bar{\Omega})$  such that

Then

$$g(\lambda) = \frac{1}{2\pi i} \int_{\partial\Omega} \frac{g(z)}{\lambda - z} dz - \frac{1}{2\pi i} \int_{\Omega} \frac{\partial_{\bar{z}} g(z)}{\lambda - z} dz d\bar{z}. \quad (3.6)$$

In particular, if  $g \in C^2(\mathbf{C})$  such that

$$\int_{\mathbf{C}} \frac{|\partial_{\bar{z}} g(z)|}{|z - \lambda|} dz d\bar{z} < \infty,$$

and  $|g(z)| \leq C \langle z \rangle^{-\varepsilon}$  with some  $\varepsilon > 0$ , then

$$g(\lambda) = -\frac{1}{2\pi i} \int_{\mathbf{C}} \frac{\partial_{\bar{z}} g(z)}{\lambda - z} dz d\bar{z}. \quad (3.7)$$

*Proof of Proposition 11.* Let  $\lambda \in \mathbf{R}$ . Then by Proposition 9,

$$|\tilde{f}(z)| \leq C \langle z \rangle^{-\varepsilon} \quad \text{and} \quad \frac{|\partial_{\bar{z}} \tilde{f}(z)|}{|z - \lambda|} \leq \frac{|\partial_{\bar{z}} \tilde{f}(z)|}{|\operatorname{Im} z|} \leq C \langle z \rangle^{-2-\varepsilon}$$

for  $z \in \mathbf{C}$ . Then by Lemma 13, we have

$$f(\lambda) = \tilde{f}(\lambda) = \frac{1}{2\pi i} \int_{\mathbf{C}} \frac{\partial_{\bar{z}} \tilde{f}(z)}{z - \lambda} dz d\bar{z} \quad \text{for } \lambda \in \mathbf{R}. \quad (3.8)$$

Now we note that

$$\frac{1}{2\pi} \int_{\mathbf{C}} |\partial_{\bar{z}} \tilde{f}| \|(A - z)^{-1}\| dz d\bar{z} \leq \frac{1}{2\pi} \int_{\mathbf{C}} \frac{|\partial_{\bar{z}} \tilde{f}|}{|\operatorname{Im} z|} dz d\bar{z} \leq C \int_{\mathbf{C}} \langle z \rangle^{-2-\varepsilon} dz d\bar{z} < \infty.$$

Hence we can use the functional calculus to obtain the formula (3.4) from (3.8). ■

Now we apply the almost analytic continuation to the  $L^p$ -boundedness of  $f(H)$ .

**Theorem 14.** Let  $m = \min\{n \in \mathbf{Z} | n > d/2 + 1\}$  and let  $\varepsilon > 0$ . Suppose  $f \in C^m(\mathbf{R})$  such that

$$\left| \left( \frac{d}{d\lambda} \right)^j f(\lambda) \right| \leq C \langle \lambda \rangle^{-\varepsilon-j} \quad \text{for } \lambda \in \mathbf{R} \quad \text{and } j = 0, 1, \dots, m. \quad (3.9)$$

Then  $f(H)$  is extended to a bounded operator in  $L^p(\mathbf{R}^d)$  for  $1 \leq p \leq \infty$ .

*Proof.* Let  $\tilde{f}(z)$  be an almost analytic continuation of  $f(\lambda)$ . Then by Proposition 9, we have

$$|\partial_{\bar{z}} \tilde{f}(z)| \leq C |\operatorname{Im} z|^{m-1} \langle z \rangle^{-\varepsilon-m} \quad \text{for } z \in \mathbf{C},$$

and by Theorem 8,

$$\|(H - z)^{-1}\|_{B(L^1)} \leq C \frac{\langle z \rangle^\beta}{|\operatorname{Im} z|^{\beta+1}} \quad \text{for } z \in \mathbf{C},$$

with  $\beta = d/2$ . Hence we have

$$\begin{aligned} & \int_{\mathbf{C}} |\partial_{\bar{z}} \tilde{f}(z)| \| (H - z)^{-1} \|_{B(L^1)} dz d\bar{z} \\ & \leq C \int_{\mathbf{C}} |\operatorname{Im} z|^{m-\beta-2} \langle z \rangle^{-(m-\beta)-\varepsilon} dz d\bar{z} < \infty \end{aligned}$$

since  $m - \beta - 2 > -1$  and  $m - \beta > 1$  by the assumption. Thus we learn

$$\|f(H)\varphi\|_{L^1} \leq C \int_{\mathbf{C}} |\partial_{\bar{z}} \tilde{f}(z)| \| (H - z)^{-1} \varphi \|_{L^1} dz d\bar{z} \leq C \|\varphi\|_{L^1}$$

for  $\varphi \in L^1 \cap L^2$ . This implies  $f(H) \in B(L^1)$  by the density argument. The claim now follows by the duality argument and complex interpolation.  $\blacksquare$

If we use almost analytic continuations for Hölder continuous functions (an idea due to E B Davies), we can improve our result slightly:

**Theorem 15.** Let  $1 \leq p \leq \infty$  and let  $\beta > d|1/p - 1/2| + 1$ . If  $f \in C^\beta(\mathbf{R})$  such that

$$|f^{(j)}(\lambda)| \leq C \langle \lambda \rangle^{-\varepsilon-j} \quad \text{for } \lambda \in \mathbf{R}, 0 \leq j \leq \beta,$$

with  $\varepsilon > 0$ , and if  $\beta \notin \mathbf{Z}$ ,

$$|f^{(\lceil \beta \rceil)}(\lambda) - f^{(\lceil \beta \rceil)}(\lambda + \mu)| \leq C \langle \lambda \rangle^{-\varepsilon-\beta} |\mu|^{\beta-\lceil \beta \rceil} \quad \text{for } \lambda, \mu \in \mathbf{R}, |\mu| \leq 1,$$

where  $\lceil \beta \rceil = \max\{n \in \mathbf{Z} | n \leq \beta\}$ . Then  $f(H)$  is extended to a bounded operator in  $L^p(\mathbf{R}^d)$ .

Let  $m$  as in Theorem 14. By the construction of  $\tilde{f}$  and the proof above, it is not hard to see

$$\|f(H)\|_{B(L^p)} \leq C \sum_{j=0}^m \sup_{\lambda} |\langle \lambda \rangle^{\varepsilon+j} f^{(j)}(\lambda)|$$

where  $C$  is independent of  $f$ . Let  $g \in C^\infty(\mathbf{R})$  such that

$$|g^{(j)}(\lambda)| \leq C \langle \lambda \rangle^{-m-\varepsilon} \quad \text{for } \lambda \in \mathbf{R} \quad \text{and } j = 0, 1, \dots, m,$$

then

$$\begin{aligned} \left| \left( \frac{d}{d\lambda} \right)^j (e^{-it\lambda} g(\lambda)) \right| &= \left| \sum_{l=0}^j \binom{j}{l} (-it)^l g^{(j-l)}(\lambda) \right| \\ &\leq C \langle t \rangle^j \langle \lambda \rangle^{-m-\varepsilon} \leq C \langle t \rangle^m \langle \lambda \rangle^{-j-\varepsilon} \end{aligned}$$

for  $\lambda \in \mathbf{R}$  and  $j = 0, 1, \dots, m$ . Hence we can apply Theorem 14 to show

$$\|e^{-itH} g(H)\|_{B(L^p)} \leq C \langle t \rangle^m \quad \text{for } t \in \mathbf{R}.$$

In fact, we can obtain estimate with respect to the order in  $t$ , at least if  $g \in C_0^\infty(\mathbf{R})$ .

**Theorem 16.** Let  $1 \leq p \leq \infty$ . Suppose  $f \in C^\infty(\mathbf{R})$  such that

$$|f^{(j)}(\lambda)| \leq C_j \langle \lambda \rangle^{-\beta-j} \quad \text{for } \lambda \in \mathbf{R}, j \geq 0,$$

with  $\beta > d|1/p - 1/2|$ . Then  $e^{-itH}f(H) \in B(L^p(\mathbf{R}^d))$  and

$$\|e^{-itH}f(H)\|_{B(L^p)} \leq C\langle t \rangle^\gamma \quad \text{for } t \in \mathbf{R}, \quad (3.10)$$

with any  $\gamma > d|1/p - 1/2|$ .

*Sketch of Proof.* For simplicity we suppose  $f \in C_0^\infty(\mathbf{R})$ . Let  $f_t(\lambda) = e^{-it\lambda}f(\lambda)$  and  $\tilde{f}_t$  be an almost analytic continuation of  $f_t$  as in Proposition 10. Then by the construction,

$$\begin{aligned} |\partial_{\bar{z}} \tilde{f}_t(z)| &\leq C_N |\text{Im } z|^N \sum_{j=0}^{N+1} \sup_{\lambda} \left| \langle \lambda \rangle^{\varepsilon+j} \left( \frac{d}{d\lambda} \right)^j (e^{-it\lambda} f(\lambda)) \right| \\ &\leq C_N \langle t \rangle^{N+1} |\text{Im } z|^N. \end{aligned}$$

On the other hand, by Lemma 6, we know

$$\|(H - z)^{-1}\|_\beta \leq C_\beta |\text{Im } z|^{-\beta-1} \quad \text{for } |z| \leq R.$$

Combining them, we have

$$\begin{aligned} \|e^{-itH}f(H)\|_\beta &= \frac{1}{2\pi} \left\| \int_{|z| \leq R} \partial_{\bar{z}} \tilde{f}_t(z) (H - z)^{-1} dz d\bar{z} \right\|_\beta \\ &\leq C_\beta \left| \int_{|z| \leq R} |\text{Im } z|^{(\beta+1)-\beta-1} \langle t \rangle^{(\beta+1)+1} dz d\bar{z} \right| = C'_\beta \langle t \rangle^{\beta+2}, \end{aligned}$$

where we set  $N = \beta + 1$ . Then we apply Theorem 5 to obtain

$$\begin{aligned} \|e^{-itH}f(H)\|_{B(\ell^1(L^2))} &\leq C \|e^{-itH}f(H)\|_\beta^{d/2\beta} \|e^{-itH}f(H)\|_{B(L^2)}^{1-d/2\beta} \\ &\leq C_\beta \langle t \rangle^{(d/2\beta)(\beta+2)} = C_\beta \langle t \rangle^{(d/2)(\beta+2)/\beta} \end{aligned}$$

if  $\beta > d/2$ . For any  $\gamma > d/2$ , we can take  $\beta$  so large that  $\gamma > (d/2)(\beta+2)/\beta$  and hence

$$\|e^{-itH}f(H)\|_{B(\ell^1(L^2))} \leq C\langle t \rangle^\gamma \quad \text{for } t \in \mathbf{R}.$$

Now we set  $g(\lambda) = (\lambda + M)^\delta f(\lambda) \in C_0^\infty(\mathbf{R}^d)$  with  $\delta > d/4$ ,  $M \gg 0$ , and apply the above result to  $g(\lambda)$ . Then since  $(\lambda + M)^{-\delta}$  is bounded from  $L^1(\mathbf{R}^d)$  to  $\ell^1(L^2)$  (Lemma 4), we learn

$$\|e^{-itH}f(H)\|_{B(L^1)} \leq C\langle t \rangle^\gamma \quad \text{for } t \in \mathbf{R}.$$

Now the result follows by the complex interpolation. ■

Actually, we can prove even slightly better estimate if the dimension is small, i.e.,  $d \leq 3$ , or  $V(x)$  is sufficiently smooth:

**Theorem 17.** Let  $d \leq 3$ . Then under the same assumptions of Theorem 16,

$$\|e^{-itH}f(H)\|_{B(L^p)} \leq C\langle t \rangle^{d|1/p - 1/2|} \quad \text{for } t \in \mathbf{R}. \quad (3.11)$$

*Sketch of Proof.* We use more direct approach than before. Instead of using the analytic continuation, we estimate  $\|e^{-itH}(H + M)^{-2}\|_\beta$  directly using the commutator

method. Namely, we first show

$$\| \langle x - n \rangle^2 e^{-itH} (H + M)^{-2} \langle x - n \rangle^{-2} \| \leq C \langle t \rangle^2$$

using the formula

$$[x, e^{-itH}] = \int_0^t [x, iH] e^{-i(t-s)H} ds.$$

This implies

$$\| e^{-itH} (H + M)^{-2} \|_2 \leq C \langle t \rangle^2,$$

and hence

$$\begin{aligned} & \| e^{-itH} (H + M)^{-2} \|_{B(\ell^1(L^2))} \\ & \leq C \| e^{-itH} (H + M)^{-2} \|_2^{d/4} \| e^{-itH} (H + M)^{-2} \|_{B(L^2)}^{1-d/4} \\ & \leq C \langle t \rangle^{d/2} \quad \text{for } t \in \mathbf{R}. \end{aligned}$$

Now the assertion follows for  $p = 1$  if  $f \in C_0^\infty(\mathbf{R})$  from this. For the general case, we again use the scaling argument as in the proof of Theorem 8. ■

## Acknowledgement

The author wants to thank Professors K B Sinha, M Krishna and The Indian Academy of Sciences for the support and hospitality during the summer school.

## References

- [AS] Aizenman M and Simon B, Brownian motion and Harnac's inequality for Schrödinger operators, *Commun. Pure. Appl. Math.* **35** (1982) 209–271
- [CT] Combes JM and Thomas L, Asymptotic behavior of eigenfunctions for multiparticle Schrödinger operators. *Commun. Math. Phys.* **34** (1973) 251–270
- [D] Davies E B, The functional calculus. Preprint, Mittag-Leffler inst., 1992/1993
- [HS] Helffer B and Sjöstrand J, Equation de Schrödinger avec champ magnétique et équation de Harper, Springer Lecture Notes in Physics **345** (1989) 118–197
- [HV] Hempel R and Voigt J, The spectrum of a Schrödinger operator in  $L^p(\mathbf{R}^n)$  is  $p$ -independent. *Commun. Math. Phys.* **104**, (1986) 243–250
- [JN1] Jensen A and Nakamura S,  $L^p$ -mapping properties of functions of Schrödinger operators and their applications to scattering theory Preprint, Aalborg Univ. 1992
- [JN2] Jensen A and Nakamura S, Mapping properties of functions of Schrödinger operators between  $L^p$ -spaces and Besov spaces Preprint, Mittag-Leffler Inst. 1992/1993
- [S] Simon B, Schrödinger semigroups. *Bull. Am. Math. Soc.* **7** (1982) 447–526



# On $N$ -body Schrödinger operators

HIROSHI ISOZAKI

Department of Mathematics, Osaka University, Toyonaka 560, Japan

**Abstract.** In these notes, we study the estimates of the resolvent or the unitary group of the  $N$ -body Schrödinger operator. The main strategy is to introduce an algebra of operators having nice commutation relations with the many-body Schrödinger operator. These estimates are applied to derive the detailed properties of the  $S$ -matrices associated with the many-body collision process.

**Keywords.**  $N$ -body problems; commutator algebra; resolvent estimates;  $S$ -matrices.

## 1. Introduction

These are the notes of my lectures delivered in the workshop on *Spectral and Inverse Spectral Theories* held at Kodaikanal in India in August 1993. The main subject is the  $N$ -body Schrödinger equation and the focus is upon the resolvent estimates and their applications to  $S$ -matrices.

A considerable progress has been made in recent years on  $N$ -body Schrödinger operators. Namely the asymptotic completeness of wave operators was proved for the 3-body short-range case by Enss [14] and for the  $N$ -body short-range case by Sigal-Soffer [48]. After these works several articles were presented by Graf [25], Tamura [53], Yafaev [57] to give a deeper insight to the mechanism of many-body Schrödinger operators. As for the long-range potentials, Enss [15] first made a breakthrough when 2-body potentials decay like  $|x|^{-\rho}$ ,  $\rho > \sqrt{3} - 1$ . This was extended by Dereziński [12] for the  $N$ -body problem. Wang [54] and Gérard [20] studied more slowly decreasing potentials.

Throughout these works lies the notion of propagation properties of quantum mechanical waves and particles in the phase space, in other words, micro-localizations of the unitary group. However the common understanding is that these micro-local estimates of the unitary group cannot be obtained by pseudo-differential operators (Ps.D.Op.'s) only but are proved by considering a framework which is bigger than the usual algebra of Ps.D.Op.'s. This is one of the most interesting features of the many-body problem.

This last point of view was introduced by Mourre [41], [42] in studying the resolvent estimates and was developed by Jensen-Mourre-Perry [39] and Jensen [37]. Further developments were then brought by Sigal-Soffer [48], [49] in the study of the estimates of the unitary group and their methods and results were refined by Dereziński [10], Skibsted [50] and Gérard [19]. With these articles in mind we introduced in our previous works [22], [23] a self-contained algebraically transparent framework to derive the resolvent estimates. Our aim here is to explain in detail the ideas in [22], [23], their by-products and their applications.

The contents of this paper are as follows:

§2. Commutator algebra; §3. Resolvent estimates (1); §4. Limiting absorption principle, revisited; §5. Resolvent estimates (2); §6. 3-body problems; §7. S-matrices in  $N$ -body problems.

In §2 we construct an algebra of operators which have a nice commutation relation with the operator

$$B = \frac{1}{2i} \left( \frac{x}{\langle x \rangle} \cdot \nabla + \nabla \cdot \frac{x}{\langle x \rangle} \right).$$

The contents of this section are deeply inspired by [48] and [10].

This algebra is used in §3 to derive the estimates of the resolvent of the  $N$ -body Schrödinger operator multiplied by functions of  $B$  or Ps.D.Op.'s. Except for the Mourre inequality our method is self-contained.

In §4 following the recent work of [33] we explain the radiation condition for the  $N$ -body Schrödinger operator and a generalization of the uniqueness theorem of Sommerfeld. As in the case of the 2-body problem, this radiation condition can be used to derive the limiting absorption principle for the resolvent. We also explain how to modify our theory to allow local singularities for the 2-body potentials.

In §5, we study more detailed estimates of the resolvent and propagation properties of the unitary group. The contents of sections 2, 3, 5 lean largely upon [22], [23]. One should also note that Wang [55] recently studied the same problem by a slightly different method. Many parts of our results overlap with his work.

In §6 we apply our results to the 3-body problem. We prove the asymptotic completeness of wave operators and state the properties of  $S$ -matrices without proof.

In §7 we shall generalize the results for the  $S$ -matrix to the  $N$ -body problem. Related topics are also mentioned and various open problems will be made clear in these sections. We excluded the proof of asymptotic completeness for the  $N$ -body case since it is long and requires different techniques. We refer the article of Soffer [52] for this problem.

Although the first big step of the asymptotic completeness has been established, a lot of problems are left open in the  $N$ -body problem. It would be our pleasure if these lecture notes help to make progress and to find new problems in this field.

## 2. Commutator algebra

### 2.1 Basic formulas

For two operators  $P$  and  $A$ , we define their multiple commutators successively by

$$\begin{aligned} ad_0(P, A) &= P, \\ ad_n(P, A) &= [ad_{n-1}(P, A), A], \quad n \geq 1. \end{aligned}$$

The fundamental formulas to calculate the commutators are as follows:

$$(ad_n(P, A))^* = (-1)^n ad_n(P^*, A^*),$$

$$ad_n(PQ, A) = \sum_{k=0}^n \binom{n}{k} ad_{n-k}(P, A) ad_k(Q, A),$$

$$[P, A^n] = \sum_{k=1}^n c_{n,k} ad_k(P, A) A^{n-k},$$

$c_{n,k}$  being constants.

## 2.2 Almost analytic extension

To represent functions of self-adjoint operators the notion of almost analytic extension is very convenient. We begin with recalling the idea due to Helffer-Sjöstrand [27].

For  $m \in \mathbf{R}$ , let  $\mathcal{F}^m$  be the set of  $C^\infty$ -functions on  $\mathbf{R}$  such that

$$|f^{(k)}(t)| \leq C_k (1 + |t|)^{m-k}, \quad \forall k \geq 0.$$

Then for  $f \in \mathcal{F}^m$  ( $m \in \mathbf{R}$ ), there exists  $F(z) \in C^\infty(\mathbf{C})$ , called an almost analytic extension of  $f$ , having the following properties:

$$F(t) = f(t), \quad t \in \mathbf{R},$$

$$|\bar{\partial}_z F(z)| \leq C_N \langle z \rangle^{m-1-N} |\operatorname{Im} z|^N, \quad \forall N \geq 0,$$

$$\operatorname{supp} F(z) \subset \{z; |\operatorname{Im} z| \leq 1 + |\operatorname{Re} z|\}. \quad (2.1)$$

Furthermore,  $\partial_t^k F(z)$  is an almost analytic extension of  $f^{(k)}(t)$  (see [19]). Let  $f \in \mathcal{F}^{-\varepsilon}$  ( $\varepsilon > 0$ ) and  $F$  be its almost analytic extension. Then the formula of Helffer-Sjöstrand tells us that for any self-adjoint operator  $A$  we have the following expression

$$f(A) = \frac{1}{2\pi i} \int_{\mathbf{C}} \bar{\partial}_z F(z) (z - A)^{-1} dz \wedge d\bar{z} \quad (2.2)$$

(see [27]). Using (2.2), one can also prove the following formula of the asymptotic expansion of the commutator: If  $f \in \mathcal{F}^m$  ( $m \in \mathbf{R}$ ) and  $A$  is self-adjoint, we have

$$[P, f(A)] = \sum_{n=1}^{N-1} (-1)^{n-1} / n! ad_n(P, A) f^{(n)}(A) + R_N, \quad (2.3)$$

$$R_N = \frac{1}{2\pi i} \int_{\mathbf{C}} \bar{\partial}_z F(z) (A - z)^{-1} ad_N(P, A) (A - z)^{-N} dz \wedge d\bar{z}. \quad (2.4)$$

Let  $\mathbf{B}$  be the set of all bounded operators. Then  $R_N \in \mathbf{B}$ , if there exists  $k$  such that  $m + k < N$  and  $ad_N(P, A)(A + i)^{-k} \in \mathbf{B}$ . This commutator expansion formula is now used in various problems in spectral and scattering theory (see [20], [24]).

## 2.3 Schrödinger operators

We consider a system of  $N$ -particles in  $\mathbf{R}^v$  with mass  $m_i$  and position  $x^i$ . Then the associated kinetic energy operator  $\tilde{H}_0$  is given by

$$M = \sum_{i=1}^N m_i,$$

$$(\mu_j)^{-1} = (m_{j+1})^{-1} + \left( \sum_{i=1}^j m_i \right)^{-1} \quad (1 \leq j \leq N-1),$$

and the *Jacobi coordinates*:

$$r_c = \sum_{i=1}^N m_i x^i / M,$$

$$r_j = x^{j+1} - \sum_{i=1}^j m_i x^i / \sum_{i=1}^j m_i \quad (1 \leq j \leq N-1).$$

Then by induction on  $N$  one can show that

$$\sum_{i=1}^N m_i (x^i)^2 = M(r_c)^2 + \sum_{i=1}^{N-1} \mu_i (r_i)^2,$$

which implies that, since the Laplacian is invariant by the linear orthogonal transformations,

$$\tilde{H}_0 = -\frac{1}{2M} \Delta_{r_c} - \sum_{i=1}^{N-1} \frac{1}{2\mu_i} \Delta_{r_i}.$$

The second term of the right-hand-side is the kinetic energy operator with the center of mass removed, which we denote by  $H_0$ . Letting  $x_i = \sqrt{2\mu_i} r_i$ , we have

$$H_0 = - \sum_{i=1}^{N-1} \Delta_{x_i}.$$

Let  $\mathcal{X}$  be defined by

$$\mathcal{X} = \{(x^1, \dots, x^N); \sum_{i=1}^N m_i x^i = 0\}. \quad (2.5)$$

Then  $-H_0$  is the Laplace-Beltrami operator on  $\mathcal{X}$  equipped with the Riemannian metric induced from  $ds^2 = 2\sum_{i=1}^N m_i (dx^i)^2$  on  $\mathbf{R}^N$ . A simple computation shows that

$$x^{j+1} - x^j = r_j - \left( \sum_{i=1}^j m_i \right)^{-1} \left( \sum_{i=1}^{j-1} m_i \right) r_{j-1},$$

which implies that  $x^i - x^j$  is a linear combination of  $x_1, \dots, x_{N-1}$ . Therefore for a real-valued function  $V_{ij}$  on  $\mathbf{R}^N$

$$H = H_0 + \sum_{i < j} V_{ij}(x^i - x^j) \quad (2.6)$$

is well-defined as an operator on  $\mathcal{X}$ , which is the (total)  $N$ -body Schrödinger operator

with the center of mass removed. We assume that each  $V_{ij}$  is a smooth function on  $\mathbf{R}^v$  and satisfies for some constant  $\rho > 0$

$$|\partial_y^m V_{ij}(y)| \leq C_m \langle y \rangle^{-m-\rho}, \quad \forall m \geq 0. \quad (2.7)$$

Here and in the sequel  $\partial_y^m$  denotes an arbitrary derivative of order  $m$  with respect to  $y$  and  $\langle y \rangle = (1 + |y|^2)^{1/2}$ .

## 2.4 Commutator algebra

As in [10] and [48], an important role is played by the self-adjoint operator  $B$  defined by

$$B = \frac{1}{2i} \left( \frac{x}{\langle x \rangle} \cdot \nabla_x + \nabla_x \cdot \frac{x}{\langle x \rangle} \right).$$

We first consider the commutation relations between  $H$ ,  $B$  and  $X = \langle x \rangle$ . Let  $L_0$  be the differential operator defined by

$$L_0 = \sum_{i=1}^{(N-1)v} x_i \frac{\partial}{\partial x_i}.$$

Let  $\mathcal{V}$  be the set of  $C^\infty$ -functions  $v$  on  $\mathcal{X}$  such that  $L_0^n v$  is bounded on  $\mathcal{X}$  for any  $n \geq 0$ . This set  $\mathcal{V}$  forms an algebra and is independent of the choice of the Jacobi coordinates.

To show an example of the element of  $\mathcal{V}$ , we take  $v \in C^\infty(\mathbf{R}^v)$  satisfying  $|\partial_y^m v(y)| \leq C_m \langle y \rangle^{-m}$ ,  $\forall m \geq 0$ . Then by choosing  $x^i - x^j$  one of the coordinates on  $\mathcal{X}$  one can see that  $v(x^i - x^j) \in \mathcal{V}$ . In particular, each two-body potential  $V_{ij}(x^i - x^j)$  belongs to  $\mathcal{V}$ .

Let  $\mathcal{V}_m = X^m \mathcal{V}$ . Let  $\mathcal{P}_{k,m}$  be the set of differential operators of order  $k$  with coefficients in  $\mathcal{V}_m$ .  $\mathcal{V}_m$  is invariant by the action of  $L_0$ , which implies that,  $[L, B] \in \mathcal{P}_{k,m-1}$  if  $L \in \mathcal{P}_{k,m}$ . We have, therefore,

*Lemma 2.1. For  $n \geq 0$ , we have*

- (1)  $ad_n(X, B) \in \mathcal{P}_{0,1-n}$ .
- (2)  $ad_n(H, B) \in \mathcal{P}_{2,-n}$ .

These commutation relations suggest to introduce the following

## DEFINITION 2.2.

$P \in \mathcal{OP}^m(X) (m \in \mathbf{R}) \Leftrightarrow X^\alpha ad_n(P, B) X^\beta \in \mathbf{B}$ , for any  $\alpha, \beta \in \mathbf{R}$  and  $n \geq 0$  such that  $\alpha + \beta = n - m$ .

We summarize the basic properties of  $\mathcal{OP}^m(X)$  in the following lemma whose proof follows easily from the definition.

- Lemma 2.3.* (1)  $P \in \mathcal{OP}^m(X) \Leftrightarrow$  There exists  $P_0 \in \mathcal{OP}^0(X)$  such that  $P = X^m P_0$ .  
 (2)  $P \in \mathcal{OP}^m(X) \Rightarrow [P, B] \in \mathcal{OP}^{m-1}(X)$ .  
 (3)  $P \in \mathcal{OP}^m(X) \Rightarrow X^k P X^l \in \mathcal{OP}^{m+k+l}(X), \quad \forall k, l \in \mathbf{R}$ .

$$(4) P \in \mathcal{OP}^m(X) \Rightarrow P^* \in \mathcal{OP}^m(X).$$

$$(5) P \in \mathcal{OP}^m(X), Q \in \mathcal{OP}^n(X) \Rightarrow PQ \in \mathcal{OP}^{m+n}(X).$$

Therefore,  $\cup_m \mathcal{OP}^m(X)$  forms an algebra which is the basic tool in our approach.

A sufficient condition for  $P$  to belong to  $\mathcal{OP}^0(X)$  is as follows. We define the operators  $S$  and  $T$  by

$$S(P) = [P, X], \quad T(P) = X[P, B]. \quad (2.8)$$

Let  $\mathbf{C}[S, T]$  be the set of all polynomials generated by  $S$  and  $T$ . Then  $P \in \mathcal{OP}^0(X)$  if  $P$  has the following property:

$$Q(P) \in \mathbf{B}, \quad \forall Q \in \mathbf{C}[S, T]. \quad (2.9)$$

This assertion is easily verified by using the following formula

$$X^n \text{ad}_n(P, B) = \sum_{k=1}^n f_{k,n}(X^{-1}) T^k(P),$$

where  $f_{k,n}$  is a polynomial,  $f_{n,n} = 1$ .

One can see a close analogy of our class  $\mathcal{OP}^0(X)$  to that of Ps.D.Op.'s, if one recalls Beal's result on the characterization of Ps.D.Op.'s by commutation relations: A bounded operator  $P$  on  $L^2(\mathbf{R}^n)$  is a Ps.D.Op. with symbol in  $S_{0,0}^0$ -class if and only if all the multiple commutators of  $P$  and  $x_i, \partial/\partial x_j$  are  $L^2(\mathbf{R}^n)$ -bounded ([7]).

We show some important examples of the elements of  $\mathcal{OP}^m(X)$ .

*Lemma 2.4.* (1)  $f(X) \in \mathcal{OP}^m(X)$  if  $f \in \mathcal{F}^m, m \in \mathbf{R}$ .

(2)  $f(H), f(B) \in \mathcal{OP}^0(X)$  if  $f \in \mathcal{F}^{-\varepsilon}, \varepsilon > 0$ .

*Proof.* (1) is obvious. To prove (2), we show that  $f(H), f(B)$  have the property (2.9). Using the notation in (2.8) and letting  $R(z) = (H - z)^{-1}$ , we have

$$S(R(z)) = -R(z)[H, X]R(z)$$

$$T(R(z)) = -R(z)X[H, B]R(z) + S(R(z))[H, B]R(z).$$

So, we have for  $Q \in \mathbf{C}[S, T]$

$$\|Q(R(z))\| \leq C_\alpha \langle z \rangle^\alpha |\text{Im } z|^{-\alpha-1} \quad (2.10)$$

for some  $\alpha > 0$ . The inequalities (2.1) and (2.10) show  $f(H) \in \mathcal{OP}^0(X)$ . The proof for  $f(B)$  is similar.  $\square$

*Lemma 2.5.* For  $f \in \mathcal{F}^m (m \in \mathbf{R})$  and  $\varphi \in C_0^\infty(\mathbf{R})$ ,  $f(B)\varphi(H) \in \mathcal{OP}^0(X)$ .

*Proof.* We take  $N > 0$  large enough, rewrite  $f(B)\varphi(H)$  as  $f(B)(B+i)^{-N}(B+i)^N\varphi(H)$  and apply the above lemma.  $\square$

It is convenient to introduce the following notation: Let  $P_n \in \mathcal{OP}^{k(n)}(X)$ ,  $k(1) > k(2) > \dots \rightarrow -\infty$ . Then an operator  $P$  is said to have the asymptotic expansion  $\Sigma_{n \geq 1} P_n$ , written as  $P \sim \Sigma_{n \geq 1} P_n$ , if and only if

$$P - \sum_{n=1}^{N-1} P_n \in \mathcal{OP}^{k(N)}(X), \quad \forall N \geq 2.$$

Lemma 2.6. Let  $P \in \mathcal{O}\mathcal{P}^m(X)$ ,  $f \in \mathcal{F}^n$ ,  $m, n \in \mathbf{R}$ . Then we have

$$[P, f(B)] \sim \sum_{k \geq 1} (-1)^{k-1} / k! \operatorname{ad}_k(P, B) f^{(k)}(B),$$

$$\operatorname{ad}_k(P, B) \in \mathcal{O}\mathcal{P}^{m-k}(X).$$

### 3. Resolvent estimates (1)

#### 3.1 Localization by $B$

The localization by the spectrum of  $B$  is a fundamental idea introduced by Sigal-Soffer [48]. We first define the following classes of functions.

#### DEFINITION 3.1

For  $a, m \in \mathbf{R}$ , let  $\mathcal{F}_{\pm}^m(a)$  be defined by

$$\mathcal{F}_{+}^m(a) = \{f \in \mathcal{F}^m; \operatorname{supp} f \subset (a, \infty)\},$$

$$\mathcal{F}_{-}^m(a) = \{f \in \mathcal{F}^m; \operatorname{supp} f \subset (-\infty, a)\}.$$

#### 3.2. Mourre estimate

By decomposing  $N$  particles into clusters and removing the intercluster potentials, we get the notion of the *cluster decomposition Hamiltonian*. The set of *thresholds*,  $\Lambda$ , is defined as the set of the eigenvalues of these cluster decomposition Hamiltonians (see e.g. [48]). Under our assumption on the 2-body potentials,  $\Lambda$  is known to be closed, countable and  $\Lambda \cap (0, \infty) = \emptyset$  (see [17]). For  $\lambda \in \sigma_{\text{cont}}(H) \cap \Lambda^c$ , we define

$$a(\lambda) = \inf\{\lambda - t; t \in \Lambda, t < \lambda\}. \quad (3.1)$$

Note that  $a(\lambda) = \lambda$  if  $\lambda > 0$ , which follows from the fact that  $\Lambda \cap (0, \infty) = \emptyset$ .

We fix  $\lambda \in \sigma_{\text{cont}}(H) \cap \sigma_p(H)^c \cap \Lambda^c$  and let  $C_0(\lambda) = a(\lambda) - \varepsilon$  for small  $\varepsilon > 0$ . Let  $\varphi \in C_0^{\infty}(\mathbf{R})$  be such that  $\varphi(t) = 1$  if  $|t - \lambda| < \delta$ ,  $\varphi(t) = 0$  if  $|t - \lambda| > 2\delta$ . Our starting point is the following Mourre type estimate [18]: For small  $\delta > 0$

$$\varphi(H) i[H, A] \varphi(H) \geq 2C_0(\lambda) \varphi(H)^2, \quad (3.2)$$

where

$$A = \frac{1}{2i} (x \cdot \nabla_x + \nabla_x \cdot x).$$

#### 3.3. Positive commutators

We are going to find an operator  $P$  whose commutator with  $H$  has certain *positivity* properties. For a small  $\varepsilon_0 > 0$ , we take  $F_0(t) \in \mathcal{F}_{-}^0(\sqrt{a(\lambda)})$  such that

$$\begin{cases} F_0(t) = 0 & \text{if } t > \sqrt{C_0(\lambda) - \varepsilon_0}, \\ F_0(t) = 1 & \text{if } t < \sqrt{C_0(\lambda) - 2\varepsilon_0}, \\ F_0(t) \geq 0, & \sqrt{F_0(t)} \in \mathcal{F}_-^0(\sqrt{a(\lambda)}), \\ F'_0(t) \leq 0, & \sqrt{-F'_0(t)} \in \mathcal{F}_-^0(\sqrt{a(\lambda)}). \end{cases}$$

For  $0 < \varepsilon_1 < \varepsilon_0$ , let  $C_1(\lambda) = \sqrt{C_0(\lambda) - \varepsilon_1}$  and define

$$F_m(t) = (C_1(\lambda) - t)^m F_0(t),$$

$$\tilde{F}_{2m+1}(t) = (C_1(\lambda) - t) F_m(t)^2.$$

For brevity, in the following arguments,  $(*)$  denotes an operator having the asymptotic expansion:

$$\sum_{n \geq 2} P_n f_n(B), \quad P_n \in \mathcal{O}\mathcal{P}^{2m+1-n}(X),$$

$$f_n \in \mathcal{F}_-^0(\sqrt{a(\lambda)}), \quad \text{supp } f_n \subset \text{supp } F_0.$$

The crucial step is the following lemma.

**Lemma 3.2.** *Let  $m > -1/2$ . With  $F_m(t)$  and  $\varphi(t)$  introduced above, we define  $P_m = X^m F_m(B) \varphi(H)$ . Then there exists a constant  $C_0 > 0$  such that*

$$- \text{Re } \varphi(H) i[H, X^{2m+1} \tilde{F}_{2m+1}(B)] \varphi(H) \geq C_0 P_m^* P_m + (*).$$

*Proof.* To calculate the commutator  $i[H, X^{2m+1} \tilde{F}_{2m+1}(B)]$  in the algebra explained in §2, we make the following device. Let  $\varphi_1(t) \in C_0^\infty(\mathbf{R})$  be such that  $\varphi_1(t) = 1$  on  $\text{supp } \varphi$ , and put  $\psi(t) = t\varphi_1(t)$ . Then

$$\begin{aligned} & \varphi(H) i[H, X^{2m+1} \tilde{F}_{2m+1}(B)] \varphi(H) \\ &= \varphi(H) i[\psi(H), X^{2m+1} \tilde{F}_{2m+1}(B)] \varphi(H) \\ &= \varphi(H) i[\psi(H), X^{2m+1}] \tilde{F}_{2m+1}(B) \varphi(H) \\ &+ \varphi(H) X^{2m+1} i[\psi(H), \tilde{F}_{2m+1}(B)] \varphi(H). \end{aligned}$$

We first show that

$$\begin{aligned} & - \text{Re } \varphi(H) X^{2m+1} i[\psi(H), \tilde{F}_{2m+1}(B)] \varphi(H) \\ & \geq (2m+1) P_m^* (2C_0(\lambda) - 2B^2 - \varepsilon_2) P_m + (*), \end{aligned} \tag{3.3}$$

$\varepsilon_2$  being a sufficiently small positive constant. In fact, we have

$$\frac{d}{dt} \tilde{F}_{2m+1}(t) = -(2m+1) F_m(t)^2 - G(t),$$

where

$$G(t) = -2(C_1(\lambda) - t)^{2m+1} F_0(t) F'_0(t).$$

Then using Lemma 2.6, we see that the left-hand side of (3.3) is written as

$$(2m+1) \text{Re } \varphi(H) X^{2m+1} i[\psi(H), B] F_{2m+1}(B)^2 \varphi(H)$$



Taking note of the relation,

$$\begin{aligned} & \varphi(H)X^{1/2}i[H, B]X^{1/2}\varphi(H) \\ &= \varphi(H)(i[H, A] - 2B^2 + K)\varphi(H), \end{aligned} \quad (3.4)$$

$K$  being a compact operator, we have

$$\begin{aligned} & \operatorname{Re} \varphi(H)X^{2m+1}i[\psi(H), B]G(B)\varphi(H) \\ &= X^m\sqrt{G(B)}\varphi(H)X^{1/2}i[H, B]X^{1/2}\varphi(H)\sqrt{G(B)}X^m + (*) \\ &\geq X^m\sqrt{G(B)}\varphi(H)(2C_0(\lambda) - 2B^2 + K)\varphi(H)\sqrt{G(B)}X^m + (*), \\ &\geq (*), \end{aligned}$$

where we have used Lemma 2.6 in the first line, (3.4) in the second line and the fact that  $-2t^2 \geq -2(C_0(\lambda) - \varepsilon_0)$  on  $\operatorname{supp} G(t)$  in the third line. We can then see that the left-hand side of (3.3) is estimated from below by

$$\begin{aligned} & (2m+1)F_m(B)X^m\varphi(H)X^{1/2}i[H, B]X^{1/2}\varphi(H)X^mF_m(B) + (*) \\ &\geq (2m+1)P_m^*(2C_0(\lambda) - 2B^2 - \varepsilon_2)P_m + (*). \end{aligned}$$

We next show that

$$\begin{aligned} & -\operatorname{Re} \varphi(H)i[\psi(H), X^{2m+1}]\tilde{F}_{2m+1}(B)\varphi(H) \\ &\geq (2m+1)P_m^*(2B^2 - 2C_1(\lambda)^2)P_m + (*). \end{aligned} \quad (3.5)$$

In fact, the left-hand side of (3.5) is written as

$$\begin{aligned} & -\operatorname{Re} \varphi(H)i[H, X^{2m+1}]\tilde{F}_{2m+1}(B)\varphi(H) + (*) \\ &= -\operatorname{Re} 2(2m+1)\varphi(H)X^{2m}B(C_1(\lambda) - B)F_m(B)^2\varphi(H) + (*). \end{aligned}$$

Since  $t \leq C_1(\lambda)$  on  $\operatorname{supp} F_m(t)$ , we have

$$-B(C_1(\lambda) - B)F_m(B)^2 \geq (B^2 - C_1(\lambda)^2)F_m(B)^2,$$

which proves (3.5).

The lemma now follows from (3.3) and (3.5). □

Let  $F_m(t)$  be as above. We call  $X^mF_m(B)$  the operator of canonical type.

**Lemma 3.3.** *Let  $m \in \mathbf{R}$ ,  $P \in \mathcal{O}\mathcal{P}^{2m}(X)$  and  $f \in \mathcal{F}_-^{2m}(\sqrt{a(\lambda)})$ . Take  $n > m, n \in \mathbf{R}$ . Then for any  $N \geq 1$ , there exist operators of canonical type  $X^{n-k/2}F_{n-k/2}(B)$  ( $k = 1, \dots, N-1$ ),  $P_N \in \mathcal{O}\mathcal{P}^{2n-N}(X)$  and a constant  $C > 0$  such that*

$$\operatorname{Re} Pf(B) \leq C \sum_{k=0}^{N-1} F_{n-k/2}(B)X^{2n-k}F_{n-k/2}(B) + P_N.$$

*Proof.* By slightly enlarging the support of  $F_n(t)$ , we see that  $\psi(t) = f(t)F_n(t)^{-2} \in$

$\mathcal{F}_-^{-\varepsilon}(\sqrt{a(\lambda)}), \varepsilon > 0$ . Then we have

$$\begin{aligned} Pf(B) &= P\psi(B)F_n(B)^2 \\ &= F_n(B)P\psi(B)F_n(B) + [P\psi(B), F_n(B)]F_n(B). \end{aligned}$$

One can then see that

$$\operatorname{Re} F_n(B)P\psi(B)F_n(B) = F_n(B)X^n P_0 X^n F_n(B),$$

where  $P_0 = P_0^* \in \mathcal{O}\mathcal{P}^0(X)$ . Therefore, for a suitable constant  $C > 0$ ,

$$\operatorname{Re} F_n(B)P\psi(B)F_n(B) \leq CF_n(B)X^{2n}F_n(B),$$

$X^n F_n(B)$  being the operator of canonical type. Since  $[P\psi(B), F_n(B)]$  has an asymptotic expansion:

$$[P\psi(B), F_n(B)] \sim \sum_{k \geq 1} P_k F_n^{(k)}(B), \quad P_k \in \mathcal{O}\mathcal{P}^{2m-k}(X),$$

we repeat the above procedure to conclude the lemma.  $\square$

### 3.4. Resolvent estimates

The results of this paper are based on the following theorem.

**Theorem 3.4.** *Let  $\lambda$  and  $\varphi$  be as above. Let  $m > -1/2$ ,  $t > 1$  and  $F \in \mathcal{F}_-^0(\sqrt{a(\lambda)})$ . Then we have*

$$X^m F(B)\varphi(H)R(\lambda + i0)X^{-m-t} \in \mathbf{B}.$$

*Proof.* We take  $\psi \in C_0^\infty(\mathbf{R})$  such that  $\psi = 1$  on  $\operatorname{supp} \varphi$ . Let  $u = \psi(H)R(\lambda + i\varepsilon)f, \varepsilon > 0$ . By Lemma 3.3, we have only to consider the case where  $X^m F(B)$  is the operator of canonical type  $X^m F_m(B)$ .

We introduce a notation here:  $Q \in \mathcal{O}\mathcal{P}_-^m(\lambda; X)$  if and only if  $Q = Pf(B)$  for some  $P \in \mathcal{O}\mathcal{P}^m(X)$  and  $f \in \mathcal{F}_-^m(\sqrt{a(\lambda)})$ .

By Lemma 3.2, we have

$$\begin{aligned} C_0 \|X^m F_m(B)\varphi(H)u\|^2 &\leq -\operatorname{Re}(i[H, Q]\varphi(H)u, \varphi(H)u) \\ &\quad + \operatorname{Re} \sum_{n=2}^{N-1} (Q_n u, u) + (Q_N u, u), \end{aligned} \quad (3.6)$$

where  $Q = X^{2m+1} \tilde{F}_{2m+1}(B)$ ,  $Q_n \in \mathcal{O}\mathcal{P}_-^{2m+1-n}(\lambda; X)$  and  $Q_N \in \mathcal{O}\mathcal{P}^{2m+1-N}(X)$ . Note that

$$\begin{aligned} &-\operatorname{Re}(i[H, Q]\varphi(H)u, \varphi(H)u) \\ &= \operatorname{Im}\{(Q\varphi(H)u, \varphi(H)f) - (Q\varphi(H)f, \varphi(H)u)\} \\ &\quad - 2\varepsilon \operatorname{Re}(Q\varphi(H)u, \varphi(H)u). \end{aligned}$$

Let  $\delta = t - 1$ . Since  $Q$  is written as

where  $P_i \in \mathcal{O}\mathcal{P}_-^{m-i-\delta}(\lambda; X)$ ,  $\tilde{P}_i \in \mathcal{O}\mathcal{P}_-^{m+t}(\lambda; X)$  and  $Q_N \in \mathcal{O}\mathcal{P}^{2m+1-N}(X)$ , we have for some  $s > 1/2$

$$|(Q\varphi(H)u, \varphi(H)f)| \leq \sum_{i=0}^{N-1} \|P_i \varphi(H)u\|^2 + C(\|X^{-s}u\|^2 + \|X^{m+t}f\|^2).$$

Here and in the sequel  $C$  denotes a constant independent of  $\varepsilon > 0$ .  $|(Q\varphi(H)f, \varphi(H)u)|$  is estimated from above in the same way. Since  $\operatorname{Re} Q$  can be written as

$$\begin{aligned} \operatorname{Re} Q &= (\tilde{F}_{2m+1}(B))^{1/2} X^{2m+1} (\tilde{F}_{2m+1}(B))^{1/2} \\ &\quad + \frac{1}{2} [(\tilde{F}_{2m+1}(B))^{1/2}, [(\tilde{F}_{2m+1}(B))^{1/2}, X^{2m+1}]], \end{aligned}$$

one can show that

$$-\operatorname{Re} \varphi(H) Q \varphi(H) \leq \sum_{i \geq 0} P_i^* P_i + Q_N,$$

with a finite number of  $P_i \in \mathcal{O}\mathcal{P}_-^{m-1/2-i}(\lambda; X)$ , and  $Q_N \in \mathcal{O}\mathcal{P}^{-N}(X)$ . Therefore for some  $s > 1/2$

$$-\operatorname{Re}(Q\varphi(H)u, \varphi(H)u) \leq \sum_{i \geq 0} \|P_i u\|^2 + C(\|X^{-s}u\|^2 + \|X^{m+t}f\|^2).$$

$\operatorname{Re}(Q_N u, u)$  in (3.6) is estimated from above similarly. So we arrive at

$$\|X^m F_m(B) \varphi(H)u\|^2 \leq \sum_{i \geq 0} \|P_i u\|^2 + C(\|X^{-s}u\|^2 + \|X^{m+t}f\|^2), \quad (3.7)$$

with a finite number of  $P_i \in \mathcal{O}\mathcal{P}_-^{m-\delta}(\lambda; X)$ . In view of Lemma 3.3, one can use (3.7) with  $m$  replaced by  $m - \delta$  to estimate  $\|P_i u\|^2$ . We repeat this procedure and finally obtain

$$\|X^m F_m(B) \varphi(H)u\|^2 \leq C(\|X^{-s}u\|^2 + \|X^{m+t}f\|^2), \quad (3.8)$$

with some  $s > 1/2$ . Now by the result of Perry–Sigal–Simon [45],  $X^{-s}R(\lambda + i0)X^{-s} \in \mathbf{B}$ , for any  $s > 1/2$ , which proves the theorem.  $\square$

Note that the inequality (3.8) was derived without using the limiting absorption principle.

By the same way as above one can show the following

**Theorem 3.5.** *Let  $\lambda$  and  $\varphi$  be as above. Let  $m > -1/2$ ,  $t > 1$  and  $F \in \mathcal{F}_+^0(-\sqrt{a(\lambda)})$ . Then we have*

$$X^m F(B) \varphi(H) R(\lambda - i0) X^{-m-t} \in \mathbf{B}.$$

#### COROLLARY 3.6

Let  $F_{\pm} \in \mathcal{F}^0$ . Suppose there exists  $\sigma \in \mathbf{R}$  such that  $|\sigma| < \sqrt{a(\lambda)}$  and  $\operatorname{supp} F_+ \subset (\sigma, \infty)$ ,  $\operatorname{supp} F_- \subset (-\infty, \sigma)$ . Then for any  $m \in \mathbf{R}$

$$X^m F_-(B) \varphi(H) R(\lambda + i0) F_+(B) X^m \in \mathbf{B}$$

*Proof.* In the proof of Theorem 3.4, we replace  $f$  by  $F_+(B)X^m f$ . We next note that we can take  $\tilde{F}_{2m+1}$  in such a way that  $\text{supp } \tilde{F}_{2m+1} \subset (-\infty, \sigma)$ , hence  $\tilde{F}_{2m+1}(B)F_+(B) = 0$ . With this in mind, we repeat the arguments in the above proof by changing the weights suitably. The details are essentially the same as in [37].  $\square$

### 3.5. Pseudo-differential operators

We shall convert the estimate in Theorem 3.4 in terms of pseudo-differential operators. For  $k > 0$  and  $a \in \mathbf{R}$ , we introduce

DEFINITION 3.7.

$\mathcal{R}_{\pm}^k(a)$  is the set of  $C^\infty$ -functions  $p(x, \xi)$  on  $\mathcal{X} \times \mathcal{X}^*$  such that

$$|\partial_x^m \partial_\xi^n p(x, \xi)| \leq C_{mn} \langle x \rangle^{-m} \langle \xi \rangle^{-k}, \quad (3.9)$$

for  $0 \leq m \leq k$ ,  $0 \leq n \leq k$  and on  $\text{supp } p(x, \xi)$

$$\inf_{x, \xi} \pm \frac{x \cdot \xi}{\langle x \rangle} > \pm a, \quad (3.10)$$

where the sign  $+$  corresponds to  $\mathcal{R}_+^k(a)$  and  $-$  to  $\mathcal{R}_-^k(a)$ .

**Theorem 3.8.** Let  $\lambda \in \sigma_{\text{ess}}(H) \cap \sigma_p(H)^c \cap \Lambda^c$  and  $a(\lambda)$  be as in (3.1). Then for any  $s > -1/2$  and  $t > 1$ , there exists  $k = k(s) > 0$  such that

$$X^s P_- R(\lambda + i0) X^{-s-t} \in \mathbf{B},$$

for any  $P_- \in \mathcal{R}_-^k(\sqrt{a(\lambda)})$ .

*Proof.* Let  $\varphi(H)$  be as above. Then by Lemma 2.4,

$$X^m (1 - \varphi(H)) R(\lambda + i0) X^{-m} \in \mathbf{B}, \quad \forall m \in \mathbf{R}.$$

Therefore to prove Theorem 3.8, we have only to consider  $\varphi(H) R(\lambda + i0)$ . For a small  $\varepsilon_0 > 0$ , we define  $C(\lambda) = \sqrt{a(\lambda)} - \varepsilon + 3\varepsilon_0$  so that  $C(\lambda) < \sqrt{a(\lambda)}$ . We take  $F_-(t) \in \mathcal{F}$  such that  $F_-(t) = 1$  if  $t < C(\lambda) - \varepsilon_0$ ,  $F_-(t) = 0$  if  $t > C(\lambda)$ . Let  $F_+(t) = 1 - F_-(t)$ . Throughout this section, we shall use the Weyl calculus of Ps.D.Op.'s.

Let  $P \in \mathcal{R}_-^k(\sqrt{a(\lambda)})$ . Then for  $s > -1/2$  one can take  $k$  large enough so that  $X^s P < B >^{-s} X^{-s} \in \mathbf{B}$ . Therefore by Theorem 3.4,

$$\begin{aligned} X^s P F_-(B) \varphi(H) R(\lambda + i0) X^{-s-t} \\ = X^s P < B >^{-s} X^{-s} \cdot X^s < B >^s F_-(B) \varphi(H) R(\lambda + i0) X^{-s-t} \in \mathbf{B} \end{aligned}$$

for  $s > -1/2$  and  $t > 1$ .

The proof of Theorem 3.8 is thus completed if we show the following assertion. For any  $s > 0$ , there exists  $k = k(s) > 0$  such that

$$X^s P F_+(B) X \in \mathbf{B}, \quad \forall P \in \mathcal{R}_-^k(\sqrt{a(\lambda)}). \quad (3.11)$$

Applying Lemma 2.6 to  $[X, F_+(B)]$ , we see that (3.11) follows from the following assertion: For any  $s > 0$ , there exists  $k = k(s) > 0$  such that

$$X^s P F_+(B) \in \mathbf{B}, \quad \forall P \in \mathcal{R}_-^k(\sqrt{a(\lambda)}). \quad (3.12)$$

Suppose (3.12) is proved for some  $s \geq 0$ . Let  $C_1(\lambda) = \sqrt{a(\lambda)} - \varepsilon + \varepsilon_0$ . Then by taking  $\varepsilon$  and  $\varepsilon_0$  small enough we have

$$\frac{x \cdot \xi}{\langle x \rangle} \leq C_1(\lambda) - \varepsilon_0$$

on  $\text{supp } p(x, \xi)$  and  $t \geq C_1(\lambda) + \varepsilon_0$  on  $\text{supp } F_+(t)$ . Let  $B_1 = B - C_1(\lambda)$  and consider

$$P(t) = e^{-tB_1} F_+(B) P^* X^{2s+1} P F_+(B) e^{-tB_1}, \quad t \geq 0.$$

Let  $b_1(x, \xi)$  be the symbol of  $B_1$ :

$$b_1(x, \xi) = \frac{x \cdot \xi}{\langle x \rangle} - C_1(\lambda).$$

Then on  $\text{supp } p(x, \xi)$ ,  $b_1(x, \xi) < -\varepsilon_0$ . Let  $P_0$  be the Ps.D.Op. with symbol

$$p_0(x, \xi) = (-b_1(x, \xi))^{1/2} p(x, \xi).$$

As is easily seen  $P_0 \in \mathcal{R}_-^{k-1}(\sqrt{a(\lambda)})$ . We now take  $k$  large enough and apply the standard symbolic calculus to obtain

$$\begin{aligned} 2P^* X^{2s+1} P_0 &= -B_1 P^* X^{2s+1} P - P^* X^{2s+1} P B_1 \\ &\quad + \text{Re} \sum_i^{\text{finite}} \tilde{P}_i^* X^{2s} P_i + Q, \end{aligned}$$

where  $P_i, \tilde{P}_i \in \mathcal{R}_-^l(\sqrt{a(\lambda)})$ ,  $l = l(k, s)$  satisfies  $l(k, s) \rightarrow \infty$  as  $k \rightarrow \infty$ , and the symbol of  $Q$  is rapidly decreasing in  $x$ . We have, therefore,

$$B_1 P^* X^{2s+1} P + P^* X^{2s+1} P B_1 \leq \text{Re} \sum_i \tilde{P}_i^* X^{2s} P_i + Q.$$

Hence by the induction hypothesis

$$\begin{aligned} -\frac{d}{dt} P(t) &\leq e^{-tB_1} F_+(B) (\text{Re} \sum_i \tilde{P}_i^* X^{2s} P_i + Q) F_+(B) e^{-tB_1} \\ &\leq C e^{-t\varepsilon_0}, \end{aligned}$$

with some constant  $C > 0$ , if  $k$  is chosen large enough. Since

$$F_+(B) P^* X^{2s+1} P F_+(B) = P(0) = - \int_0^\infty \frac{d}{dt} P(t) dt,$$

one can see that  $X^{s+1/2} P F_+(B) \in \mathbf{B}$ , which completes the proof of Theorem 3.8.  $\square$

## 4. Limiting absorption principle, revisited

### 4.1. Radiation condition

For  $s \in \mathbf{R}$ , let  $L^{2,s} = L^{2,s}(\mathbf{R}^n)$  be the weighted Hilbert space defined by

$$u \in L^{2,s} \Leftrightarrow \|u\|_s = \|\langle x \rangle^s u(x)\|_{L^2} < \infty.$$

For two Banach spaces  $\mathcal{H}_1$  and  $\mathcal{H}_2$ , let  $\mathbf{B}(\mathcal{H}_1; \mathcal{H}_2)$  denote the set of all bounded operators from  $\mathcal{H}_1$  to  $\mathcal{H}_2$ .

By the limiting absorption principle, we mean the existence of the boundary values of the resolvent  $R(\lambda \pm i0) \in \mathbf{B}(L^{2,s}; L^{2,-s})$  for some  $s > 0$ , where  $R(z) = (H - z)^{-1}$ ,  $\lambda \in \sigma_{\text{cont}}(H)$ . There are several approaches to this problem.

The first one is the classical partial differential theoretical approach due to Eidus [13], Jäger [36] and Ikebe-Saito [29] for example. This method mainly deals with the 2-body Schrödinger operator, derives some a-priori estimates for the equation  $(-\Delta + V - z)u = f$  by integration by parts and reduces the existence of the boundary value  $(-\Delta + V - \lambda \mp i0)^{-1}$  to the following uniqueness theorem.

**Theorem 4.1.** *Let  $H = -\Delta + V$  be the Schrödinger operator on  $\mathbf{R}^n$ , where  $V(x)$  is a real function satisfying*

$$|\partial_x^m V(x)| \leq C \langle x \rangle^{-m-\rho}, \quad m = 0, 1, \quad \rho > 0.$$

*Let  $0 \leq \alpha < 1/2 < s \leq 1$ . Suppose that  $u \in L^{2,-s}$  satisfies  $(H - \lambda)u = 0$ ,  $\lambda > 0$  and that one of the following conditions is satisfied:*

$$\left( \frac{\partial}{\partial r} \mp i\sqrt{\lambda} \right) u \in L^{2,-\alpha}. \quad (4.1)$$

*Then  $u = 0$ .*

Furthermore if  $f \in L^{2,s}$  for some  $s > 1/2$ ,  $u_{\pm} = R(\lambda \pm i0)f$  satisfies (4.1). More precisely  $((\partial/\partial r) - i\sqrt{\lambda})u_{+} \in L^{2,-\alpha}$ , and  $((\partial/\partial r) + i\sqrt{\lambda})u_{-} \in L^{2,-\alpha}$ . Therefore, the condition (4.1) characterizes  $R(\lambda \pm i0)$  in the sense that  $u_{\pm} \in L^{2,-s}$  satisfying  $(H - \lambda)u = f$ ,  $f \in L^{2,s}$ , is written as  $u_{\pm} = R(\lambda \pm i0)f$  if and only if  $u_{\pm}$  satisfies the *radiation condition* (4.1).

The second approach is of functional theoretic character due to Agmon [1] and Kuroda [40] based on the Fourier analysis. It is applicable to a wide range of equations although limited to short-range perturbations. Agmon further invented a method applicable to the long-range case containing various deep insights to the micro-local calculus for the spectral theory (see [2], [28]).

The third approach based on the commutator technique was introduced by Mourre [41]. This is a purely abstract method, includes in particular the many-body Schrödinger operators (see also [45]) and has now become a fundamental tool in spectral and scattering theory.

However, since Mourre's approach asserts only the existence of the boundary values of the resolvent  $R(\lambda \pm i0)$  and does not give its characterization, it seems to be worthwhile to investigate this problem. Let us return to the  $N$ -body Schrödinger operator introduced in §3. The following theorem was proved in [33].

**Theorem 4.2.** *Let  $0 \leq \alpha < 1/2 < s \leq 1$  and  $\lambda \in \sigma_{\text{cont}}(H) \cap \sigma_p(H)^c \cap \Lambda^c$ . Suppose that*

$u \in L^{2,-s}$  satisfies  $(H - \lambda)u = 0$  and that there exists an  $\varepsilon > 0$  such that

$$F(B)u \in L^{2,-\alpha} \quad (4.2)$$

either  $\forall F \in \mathcal{F}_-^0(\varepsilon)$  or  $\forall F \in \mathcal{F}_+^0(-\varepsilon)$ . Then  $u = 0$ .

Here we note that for  $F \in \mathcal{F}^0$ ,  $F(B) \in \mathbf{B}(L^{2,s}; L^{2,s})$  for any  $s \in \mathbf{R}$ . In fact, this follows easily from (2.3) and the formula

$$X^s F(B) X^{-s} = F(B) - X^s [X^{-s}, F(B)].$$

Therefore  $F(B)u$  makes sense.

One should note that the condition (4.2) is weaker than (4.1). In fact we have

**Lemma 4.3.** *Let  $\lambda > 0$ . Suppose that  $(B - \sqrt{\lambda})u \in L^{2,-\alpha}$ . Then  $F(B)u \in L^{2,-\alpha}$  for any  $F \in \mathcal{F}_-^0(\varepsilon)$ ,  $0 < \varepsilon < \sqrt{\lambda}$ .*

The proof is easy if one notes that  $F_1(t) = F(t)(t - \sqrt{\lambda})^{-1} \in \mathcal{F}_-^{-1}(\varepsilon)$  and  $F(B) = F_1(B)(B - \sqrt{\lambda})$ .

In view of Theorems 3.4, 4.1, 4.2 and Lemma 4.3, one can see that the condition (4.2) characterizes  $R(\lambda \pm i0)$  and is a generalization of the radiation condition of Sommerfeld.

#### 4.2. Limiting absorption principle

When compared with the 2-body problem, it is not surprising that the above uniqueness theorem (Theorem 4.2), which is independent of the resolvent estimates, serves as a key-step towards the limiting absorption principle. In this subsection we shall give an alternative proof of it.

Let  $I$  be a compact interval in  $\sigma_{\text{cont}}(H) \cap \sigma_p(H)^c \cap \Lambda^c$  and

$$J = \{z \in \mathbf{C}; \operatorname{Re} z \in I, 0 < \operatorname{Im} z < 1\}.$$

We consider  $u(z) = R(z)f$ ,  $z \in J$ . We take  $\rho \in C^\infty(\mathbf{R})$  such that  $\rho(k) = 1$  for  $k \geq 2$ ,  $\rho(k) = 0$  for  $k \leq 1$  and put  $\rho_t(k) = \rho(k/t)$ .

**Lemma 4.4.** *Let  $1/2 < s < 1$ . Then there exist constants  $1/2 < \beta < s$ ,  $\varepsilon > 0$  and  $C > 0$  such that*

$$\|\rho_t(X)u(z)\|_{-\beta} \leq Ct^{-\varepsilon}(\|u(z)\|_{-\beta} + \|f\|_s),$$

for  $z \in J$ ,  $t > 1$ .

*Proof.* We first take  $F_\pm \in C^\infty(\mathbf{R})$  such that  $F_+^2 + F_-^2 = 1$ ,  $F_+(k) = 1$  for  $k > \varepsilon_0$ ,  $F_+(k) = 0$  for  $k < \varepsilon_0/2$ ,  $\varepsilon_0$  being a sufficiently small positive constant.

Now we recall that in (3.8) we have already obtained the following uniform estimate

$$\|F_-(B)u(z)\|_{-\alpha} \leq C(\|u(z)\|_{-\beta} + \|f\|_s) \quad \text{for } z \in J, \quad (4.3)$$

where  $0 < \alpha < 1/2 < \beta < s$ . With this  $\beta$ , we define  $\chi_t(k)$  by

$$\chi_t(k) = \int_k^\infty x^{-2\beta} \rho_t(x)^2 dx.$$

$$|\chi_t(k)| \leq Ct^{1-2\beta}. \quad (4.4)$$

A simple calculation shows that

$$-i([H, \chi_t(X)]u, u) = 2(BX^{-\beta}\rho_t(X)u, X^{-\beta}\rho_t(X)u).$$

On the other hand, using the equation  $(H - z)u = f$ , we have

$$-i([H, \chi_t(X)]u, u) = -2 \operatorname{Im} z(\chi_t u, u) + 2 \operatorname{Im}(\chi_t u, f).$$

Since  $\operatorname{Im} z > 0$ , we have

$$(BX^{-\beta}\rho_t(X)u, X^{-\beta}\rho_t(X)u) \leq |(\chi_t u, f)|. \quad (4.5)$$

Since

$$F_-(B)X^{-\beta}\rho_t u = X^{-\beta}\rho_t F_-(B)u + [F_-(B), X^{-\beta}\rho_t]u,$$

in view of (4.3) one can easily see that for some  $\varepsilon > 0$ ,

$$\|F_-(B)X^{-\beta}\rho_t u\| \leq Ct^{-\varepsilon}(\|u(z)\|_{-\beta} + \|f\|_s). \quad (4.6)$$

On the other hand, since there exists some  $C_0 > 0$  such that  $k > C_0$  on  $\operatorname{supp} F_+(k)$ , we have using (4.5),

$$\begin{aligned} C_0 \|F_+(B)X^{-\beta}\rho_t u\|^2 &\leq (BF_+(B)^2 X^{-\beta}\rho_t u, X^{-\beta}\rho_t u) \\ &\leq |(\chi_t u, f)| - (BF_-(B)^2 X^{-\beta}\rho_t u, X^{-\beta}\rho_t u). \end{aligned}$$

By (4.4),  $|(\chi_t u, f)| \leq Ct^{1-2\beta}(\|u(z)\|_{-\beta}^2 + \|f\|_s^2)$ . To estimate the second term, we take  $\varphi \in C_0^\infty(\mathbf{R})$  such that  $\varphi = 1$  in a neighborhood of  $I$  and split  $u(z)$  into two parts:  $\varphi(H)u(z) + (1 - \varphi(H))u(z) \equiv u_1(z) + u_2(z)$ . Then  $u_2(z)$  is uniformly bounded in  $L^2$ . So, we have only to consider the case where  $u(z)$  is replaced by  $u_1(z) = \varphi(H)u(z)$ . However it is rather easy to see that  $BF_-(B)\varphi(H)u(z)$  obeys the estimate (4.3). So, one can estimate the second term from above by  $t^{-2\varepsilon}(\|u(z)\|_{-\beta}^2 + \|f\|_s^2)$ . Hence we obtain

$$\|F_+(B)X^{-\beta}\rho_t u\| \leq Ct^{-\varepsilon}(\|u(z)\|_{-\beta} + \|f\|_s). \quad (4.7)$$

The inequalities (4.6) and (4.7) prove the lemma.  $\square$

We can now prove the following uniform estimate.

**Theorem 4.5.** *For  $s > 1/2$ , we take  $\beta$  specified in Lemma 4.4. Then there exists a constant  $C > 0$  such that*

$$\|R(z)f\|_{-\beta} \leq C\|f\|_s, \quad \forall z \in J.$$

*Proof.* Suppose this is not true. Then there exist  $z_n \in J, f_n \in L^{2,s}$  such that  $z_n \rightarrow \lambda \in I$  and  $u_n = R(z_n)f_n$  satisfies

$$\|u_n\|_{-\beta} = 1, \|f_n\|_s \rightarrow 0.$$



By Lemma 4.4 and Rellich's compactness theorem,  $\{u_n\}_{n=1}^\infty$  contains a subsequence which converges to some  $u$  in  $L^{2,-\beta}$ . Hence  $\|u\|_{-\beta} = 1$ ,  $(H - \lambda)u = 0$ . By (4.3) one can also see that  $u$  satisfies (4.2). According to Theorem 4.2,  $u = 0$ . This is a contradiction.  $\square$

**Theorem 4.6.** (*Limiting Absorption Principle*). Let  $s > 1/2$ . Then there exists a constant  $1/2 < \beta < s$  such that as a bounded operator in  $\mathbf{B}(L^{2,s}; L^{2,-\beta})$ ,  $R(z)$  is uniformly continuous with respect to  $z \in J$ .

*Proof.* Let  $\varphi \in C_0^\infty(\mathbf{R})$  be such that  $\varphi(k) = 1$  if  $k \in I$ . We have only to consider  $\varphi(H)R(z)$ . Suppose this is not uniformly continuous. Then there exist  $\varepsilon > 0$ ,  $z_n, z'_n \in J$  and  $f_n \in L^{2,s}$  such that

$$\begin{aligned} \|f_n\|_s &= 1, \quad z_n, z'_n \rightarrow \lambda \in I, \\ \|(R(z_n) - R(z'_n))\varphi(H)f_n\|_{-\beta} &\geq \varepsilon. \end{aligned}$$

Let  $u_n = R(z_n)\varphi(H)f_n$  and  $v_n = R(z'_n)\varphi(H)f_n$ . Take  $1/2 < s' < s$ . One can assume that  $\varphi(H)f_n$  converges in  $L^{2,s'}$  by compactness. Let  $w_n = u_n - v_n$ . Then  $\|w_n\|_{-\beta} \geq \varepsilon$ . Using Lemma 4.4 and Theorem 4.5, one can assume that  $w_n$  converges to some  $w$  in  $L^{2,-\beta}$ . So,  $\|w\|_{-\beta} \geq \varepsilon$ . On the other hand, it is easy to show that  $(H - \lambda)w = 0$  and  $w$  satisfies the radiation condition (4.2). So,  $w = 0$ , which is a contradiction.  $\square$

#### COROLLARY 4.7

Let  $s > 1/2$ . Then  $R(z)$  is uniformly continuous in  $\mathbf{B}(L^{2,s}; L^{2,-s})$  with respect to  $z \in J$ , and there exists a constant  $C > 0$  such that

$$\|R(z)f\|_{-s} \leq C\|f\|_s, \quad z \in J.$$

*Proof.* For a given  $s > 1/2$ , we choose  $\beta, s'$  such that  $1/2 < \beta < s' < s$ . and consider

$$X^{-s}R(z)X^{-s} = X^{-(s-\beta)} \cdot X^{-\beta}R(z)X^{-s'} \cdot X^{-(s-s')}.$$

Applying Theorems 4.6 and 4.7 to  $X^{-\beta}R(z)X^{-s'}$ , we conclude this corollary.  $\square$

#### 4.3. Singular potentials

Contrary to the 2-body problem, it is not so obvious in the  $N$ -body case to allow local singularities for the 2-body potentials, since these singularities spread to infinity in the  $N$ -body problem. The remedy comes from the clever choice of the vector field constructed by Graf [25], which is not only powerful to prove the asymptotic completeness (see also [57]) but also to prove the dilation analyticity (see [21]).

Suppose that each 2-body potentials are split into two parts:  $V_{ij} = V_{ij}^{(1)} + V_{ij}^{(2)}$ , where  $V_{ij}^{(1)}(y)$  is a smooth function satisfying the assumption (2.7),  $V_{ij}^{(2)}(y)$  is compactly supported and  $V_{ij}^{(2)}(y)(-\Delta_y + 1)^{-1}$  is compact on  $L^2(\mathbf{R}_y^v)$ .

Let  $\lambda \in \sigma_{\text{cont}}(H) \cap \sigma_p(H)^c \cap \Lambda^c$ . Then there exists a smooth function  $r(x)$  on  $\mathcal{X}$  having the following properties:

$$|\partial_x^m(x \cdot \nabla)^k(r(x)^2 - |x|^2)| \leq C_{m,k}, \quad \forall m, k. \quad (4.8)$$

Define  $\omega(x) = r(x)\nabla r(x)$  and

$$\tilde{A} = \frac{1}{2i}(\omega(x) \cdot \nabla + \nabla \cdot \omega(x)). \quad (4.9)$$

Then  $\omega(x)$  satisfies

$$|\partial_x^m (x \cdot \nabla)^k (\omega(x) - x)| \leq C_{m,k}, \quad \forall m, k, \quad (4.10)$$

and also

$$[V_{ij}^{(2)}(x^i - x^j), \tilde{A}] = 0. \quad (4.11)$$

Moreover in the same way as in (3.2) the following Mourre estimate holds:

$$\varphi(H)i[H, \tilde{A}]\varphi(H) \geq 2C_0(\lambda)\varphi(H)^2. \quad (4.12)$$

(4.11) follows from the remarkable fact that the vector field  $\omega(x)$  is parallel to  $\mathcal{X}^{(ij)} = \{x \in \mathcal{X}; x^i = x^j\}$  near this subspace. For the proof see [50], [51] and also [11].

With this vector field  $\omega(x)$ , we define the operator  $B$  by

$$B = \frac{1}{2i}((\nabla r)(x) \cdot \nabla + \nabla \cdot (\nabla r)(x)). \quad (4.13)$$

The results of this paper are then extended to these singular potentials by using this operator in the definition of § 2. We refer [23] for the details.

## 5. Resolvent estimates (2)

In this section we summarize some by-products of the resolvent estimates proved in § 3.

### 5.1. Disintegration and resolvent estimates

Quantum mechanical particles in the scattering state are, in the remote future, expected to disintegrate into several clusters moving independently each other. We prove a counterpart of this disintegration for the resolvent estimates.

We first introduce some standard notations. Let  $a = \{C_1, \dots, C_k\}$  ( $1 \leq k \leq N$ ) be a cluster decomposition. We denote by  $\#a$  the number of clusters in  $a$ . For a pair of particles  $(ij)$ ,  $(ij) \subset a$  means that the pairs  $i$  and  $j$  live in the same cluster in  $a$  and  $(ij) \not\subset a$  its negation. Let

$$\mathcal{X}_a = \{x \in \mathcal{X}; x^i = x^j \text{ if } (ij) \subset a\} \quad (5.1)$$

and  $\mathcal{X}^a$  be its orthogonal complement in  $\mathcal{X}$ . We decompose  $x \in \mathcal{X} = \mathcal{X}^a \oplus \mathcal{X}_a$  into  $x = (x^a, x_a)$  accordingly. Physically,  $x^a$  represent the coordinates within the cluster and  $x_a$  between the clusters. We decompose  $H_0 = -\Delta$  into two parts:  $H_0 = -\Delta_{x^a} - \Delta_{x_a} \equiv T^a + T_a$  and put

$$H_a = H_0 + V^a, \quad V^a = \sum_{(ij) \subset a} V_{ij}(x^i - x^j), \quad (5.2)$$

$$H^a = T^a + V^a. \quad (5.3)$$

Let  $\text{Loc}(a)$  be the set of smooth functions  $f(x)$  on  $\mathcal{X}$  such that

$$|\partial_x^m f(x)| \leq C_m \langle x \rangle^{-m}, \quad \forall m \geq 0, \quad (5.4)$$

and there exist  $0 < \varepsilon_1, \varepsilon_2 < 1$  such that on  $\text{supp} f$

$$|x| \geq 1, |x^a| \leq \varepsilon_1 |x|, |x^i - x^j| \geq \varepsilon_2 |x| \quad \text{if } (ij) \neq a. \quad (5.5)$$

On the support of the functions in  $\text{Loc}(a)$ ,  $H$  is approximated by  $H_a$  in the following sense.

*Lemma 5.1.* Let  $\varphi \in C_0^\infty(\mathbf{R})$  and  $J_a \in \text{Loc}(a)$ . Then we have for any  $N > 0$ ,

$$J_a(\varphi(H) - \varphi(H_a)) = \sum_{k=1}^{N-1} X^{-k\rho} P_k \varphi_k(H_a) + R_N,$$

where  $P_k$  is a differential operator with coefficients in  $\text{Loc}(a)$ ,  $\varphi_k \in C_0^\infty(\mathbf{R})$ ,  $\text{supp } \varphi_k \subset \text{supp } \varphi$  and  $R_N \in \mathcal{O}\mathcal{P}^{-N\rho}(X)$ .

*Proof.* Using (2.2) and the resolvent formula, we have

$$J_a(\varphi(H) - \varphi(H_a)) = \frac{1}{2\pi i} \int_C \bar{\partial}_z \tilde{\varphi}(z) J_a R(z) (V - V^a) R_a(z) dz \wedge d\bar{z}, \quad (5.6)$$

where  $\tilde{\varphi}(z)$  is an almost analytic extension of  $\varphi$ ,  $R(z) = (H - z)^{-1}$  and  $R_a(z) = (H_a - z)^{-1}$ . We take  $J'_a \in \text{Loc}(a)$  such that  $J_a J'_a = J_a$  and commute  $J'_a$  and  $R(z)$ . Noting  $X^\rho J'_a (V - V^a) \in \text{Loc}(a)$ , we commute  $J'_a (V - V^a)$  and  $R(z)$  and replace  $R(z)$  by  $R_a(z)$  by using the resolvent equation. We then find  $X^{-\rho} P_1 \varphi^{(1)}(H_a)$  as a leading term of the right-hand side of (5.6). We repeat this procedure to obtain the lemma.  $\square$

For a self-adjoint operator  $A$  and a constant  $C$ ,  $F(A < C)$  means the operator  $F_0(A)$  where  $F_0 \in \mathcal{F}_-^0(C)$  and for a constant  $\varepsilon > 0$ ,  $F_0(t) = 1$  if  $t < C - \varepsilon$ .

*Lemma 5.2.* Let  $A, B$  be self-adjoint operators such that  $A - B$  is bounded self-adjoint. Then as  $R \rightarrow \infty$ ,

$$\|F(A < 1)F(B > R)\| = O(R^{-1/2}).$$

*Proof.* We put

$$G(t, R) = e^{-tB} F(B > R) F(A < 1)^2 F(B > R) e^{-tB}.$$

Then a simple manipulation shows that for a constant  $C > 0$

$$\begin{aligned} -\frac{d}{dt} G(t, R) &= 2 \operatorname{Re} e^{-tB} F(B > R) (B - A) F(A < 1)^2 F(B > R) e^{-tB} \\ &\quad + 2 \operatorname{Re} e^{-tB} F(B > R) A F(A < 1)^2 F(B > R) e^{-tB} \\ &\leq C e^{-2tR}. \end{aligned}$$

Integrating this inequality and noting  $G(0, R) = F(B > R) F(A < 1)^2 F(B > R)$ , we obtain the Lemma.  $\square$

Similarly to Definition 3.6, for a cluster decomposition  $a$ , we introduce the following class of symbols:

### DEFINITION 5.3

For  $k > 0$  and  $\tau \in \mathbf{R}$ , let  $\mathcal{R}_\pm^k(a, \tau)$  be the set of  $C^\infty$ -functions  $p(x, \xi_a)$  on  $\mathcal{X} \times \mathcal{X}_a^*$  such that

$$|\partial_x^m \partial_{\xi_a}^n P(x, \xi_a)| \leq C_{mn} \langle x \rangle^{-m} \langle \xi_a \rangle^{-k}, \quad (5.7)$$

for  $0 \leq m \leq k$ ,  $0 \leq n \leq k$  and on  $\text{supp } p(x, \xi_a)$

$$\inf_{x_a, \xi_a} \pm \frac{x_a \cdot \xi_a}{\langle x_a \rangle} > \pm \tau, \quad (5.8)$$

where the sign '+' corresponds to  $\mathcal{R}_+^k(a, \tau)$  and - to  $\mathcal{R}_-^k(a, \tau)$ .

We prepare one more notation. For  $\varepsilon > 0$ , let  $\text{Loc}(a, \varepsilon)$  be the set of functions  $f$  in  $\text{Loc}(a)$  such that  $|x^a| \leq \varepsilon |x|$  on  $\text{supp } f$ .

**Theorem 5.4.** Let  $\lambda \in \sigma_{\text{cont}}(H) \cap \sigma_p(H)^c \cap \Lambda^c$  and  $a(\lambda)$  be as in (3.1). Let  $a$  be a cluster decomposition. Take  $\varepsilon > 0$  arbitrarily. Then there exists  $\varepsilon' > 0$  such that for any  $m > -1/2$ ,  $t > 1$ , there exists  $k > 0$  such that

$$X^m J_a P_a R(\lambda + i0) X^{-m-t} \in \mathbf{B},$$

for any  $P_a \in \mathcal{R}_-^k(a, \sqrt{a(\lambda)} - \varepsilon)$  and  $J_a \in \text{Loc}(a, \varepsilon')$ .

*Proof.* In view of Theorem 3.4, we have only to show the following assertion: Let  $\varphi \in C_0^\infty(\mathbf{R})$ . Then for any  $\varepsilon > 0$  there exists  $\varepsilon' > 0$  having the following property. For any  $m > -1/2$  and  $t > 1$ , there is  $k > 0$  such that for any  $P_a \in \mathcal{R}_-^k(a, \sqrt{a(\lambda)} - \varepsilon)$ ,  $J_a \in \text{Loc}(a, \varepsilon')$  and  $N > 0$ ,  $\varphi(H) X^m J_a P_a$  is written as

$$\varphi(H) X^m J_a P_a = \sum_{\text{finite}} Q_m P + R_N, \quad (5.9)$$

where  $Q_m \in \mathcal{OP}^m(X)$ ,  $P \in \mathcal{R}_-^l(\sqrt{a(\lambda)})$ ,  $l = l(k) \rightarrow \infty$  as  $k \rightarrow \infty$ , and  $R_N \in \mathcal{OP}^{-N}(X)$ .

We first take  $\psi \in C_0^\infty(\mathbf{R})$  such that  $\psi\varphi = \varphi$  and put  $K = F(H_0 > R)\psi(H)$ . Then  $\varphi(H) = (K + F(H_0 < R))\varphi(H)$ . By Lemma 5.2, one can take  $R$  large enough so that  $\|K\| < 1/2$ . Then  $\varphi(H) = (1 - K)^{-1} F(H_0 < R)\varphi(H)$ . By the proof of Lemma 2.4,  $K$  satisfies (2.9), hence so does  $L = (1 - K)^{-1}$ . Therefore  $L \in \mathcal{OP}^0(X)$ , which shows that we have only to study  $F(H_0 < R)\varphi(H) X^m J_a P_a$ . We next commute  $\varphi(H)$  and  $X^m J_a P_a$  by using (2.2) and noting that  $[V_{ij}, P_a] = 0$  if  $(ij) \subset a$ . This reduces the problem to investigate  $F(H_0 < R) X^m J_a P_a$ .

On the support of the symbol of this Ps.D.Op.  $F(H_0 < R) X^m J_a P_a$ ,  $|\xi^a| < C$  for some constant  $C > 0$ . Therefore we have

$$\frac{x \cdot \xi}{\langle x \rangle} \leq \frac{x_a \cdot \xi_a}{\langle x_a \rangle} + C\varepsilon' < \sqrt{a(\lambda)} - \varepsilon + C\varepsilon'.$$

Choosing  $\varepsilon'$  small enough we see that  $X^{-m} F(H_0 < R) X^m J_a P_a \in \mathcal{R}_-^k(\sqrt{a(\lambda)})$  modulo a term rapidly decreasing in  $x$ . This proves the theorem.  $\square$

## 5.2. Derivatives of the resolvent

In view of the relation

$$\left(\frac{d}{dz}\right)^n R(z) = n! R(z)^{n+1},$$

one can easily obtain the estimates of the derivatives of the resolvent by inserting  $I = F_+(B) + F_-(B)$  suitably and using Theorems 3.4, 3.5, Corollary 3.6 and their variants. We omit the detailed proof, since it is essentially the same as the one given by Jensen [37] or [30].

**Theorem 5.5.** *Let  $\lambda$  be as in §3. Then for any integer  $n \geq 0$  and  $s' > s > n - 1/2$ , there exists  $k > 0$  such that*

$$X^{s-n} P_- R(\lambda + i0)^n X^{-s'} \in \mathbf{B}$$

for any  $P_- \in \mathcal{R}_-^k(\sqrt{a(\lambda)})$ .

## 5.3. Estimates of the unitary group

We derive decay estimates of the unitary group by using Theorem 5.5.

**Theorem 5.6.** *Let  $I$  be a compact interval in  $\sigma_{\text{cont}}(H) \cap \sigma_p(H)^p \cap \Lambda^c$  and take  $\varphi \in C_0^\infty(I)$ . Let  $\sigma = \inf_{\lambda \in I} \sqrt{a(\lambda)}$ . Then for any  $m > 0$ ,  $s > -1/2$  and  $s' > s + m$ , there exists  $k > 0$  such that*

$$\|X^s P_- e^{-itH} \varphi(H) X^{-s'}\| \leq C(1+t)^{-m}$$

for any  $t > 0$  and  $P_- \in \mathcal{R}_-^k(\sigma)$ .

*Proof.* Let  $E'(\lambda) = 1/2\pi i(R(\lambda + i0) - R(\lambda - i0))$ . Then we have by integration by parts

$$\begin{aligned} e^{-itH} \varphi(H) &= \int_{-\infty}^{\infty} e^{-it\lambda} E'(\lambda) \varphi(H) d\lambda \\ &= (it)^{-n} \int_{-\infty}^{\infty} e^{-it\lambda} \left(\frac{d}{d\lambda}\right)^n E'(\lambda) \varphi(H) d\lambda. \end{aligned}$$

Note that  $(d/d\lambda)^n E'(\lambda) \varphi(H)$  exists as a bounded operator in  $\mathbf{B}(L^{2,\alpha}; L^{2,-\alpha})$  with  $\alpha > n - 1/2$  (see e.g. [37]).

By the Paley-Wiener theorem we have for  $t > 0$

$$\int_{-\infty}^{\infty} e^{-it\lambda} R(\lambda - i0)^n \varphi(H) d\lambda = 0.$$

We have, therefore, for  $t > 0$

$$P_- e^{-itH} \varphi(H) = (it)^{-n} n! \int_{-\infty}^{\infty} e^{-it\lambda} P_- R(\lambda + i0)^n \varphi(H) d\lambda.$$

Theorem 5.6 then follows from Theorem 5.5 and an interpolation.  $\square$

From Theorem 5.6 follows the optimal decay estimate in the *free region*. Let

$$\mathcal{Y}_0 = \{x \in \mathcal{X}; x^i \neq x^j \quad \forall i \neq j\}. \quad (5.10)$$

**Theorem 5.7.** *Let  $I$  be a compact interval in  $(0, \infty)$  and take  $\varphi \in C_0^\infty(I)$ . Let  $\Gamma$  be a closed cone in  $\mathcal{Y}_0$ . Let  $P$  be a Ps.D.Op. with symbol  $p(x, \xi)$  satisfying*

$$|\partial_x^m \partial_\xi^n p(x, \xi)| \leq C_{mn} \langle x \rangle^{-m} \langle \xi \rangle^{-n}, \quad \forall m, n \geq 0. \quad (5.11)$$

$$\text{supp}_x p(x, \xi) \subset \Gamma, \quad \forall \xi \in \mathcal{X}^*, \quad (5.12)$$

and there exists  $\varepsilon > 0$  such that

$$\frac{x}{\langle x \rangle} \cdot \xi < (1 - \varepsilon) |\xi| \quad (5.13)$$

on  $\text{supp } p(x, \xi)$ . Then if  $|I|$  is sufficiently small we have for any  $m > 0, s > -1/2$  and  $s' > s + m$

$$\|X^s P e^{-itH} \varphi(H) X^{-s'}\| \leq C(1+t)^{-m}$$

for any  $t > 0$ .

*Proof.* By Lemma 5.1,  $P\varphi(H)$  has the following asymptotic expansion

$$P\varphi(H) \sim \sum_{n=1}^{\infty} X^{-n\rho} P_n,$$

where  $P_n$  is a Ps.D.Op. with symbol  $p_n(x, \xi)$  having the same property as  $p(x, \xi)$ , moreover if  $(x, \xi) \in \text{supp } p_n, |\xi|^2 \in I$ . Therefore if  $|I|$  is sufficiently small

$$\frac{x}{\langle x \rangle} \cdot \xi < \inf_{\lambda \in I} \sqrt{a(\lambda)}$$

on  $\text{supp } p_n$ . The theorem then follows from Theorem 5.6.  $\square$

Passing to the Laplace transform we obtain

### COROLLARY 5.8

Let  $\lambda > 0$  and  $P$  be as in Theorem 5.7. Then for any  $s > -1/2$  and  $t > 1$  we have

$$X^s P R(\lambda + i0) X^{-s-t} \in \mathbf{B}.$$

Of course one can prove Corollary 5.8 directly from Theorem 5.4.

In the same way as in Theorem 5.6, using Theorem 5.4, one can obtain the decay estimate of the unitary group with  $P_-$  replaced by  $J_a P_a$ , where  $J_a \in \text{Loc}(a, \varepsilon')$ ,  $P_a \in \mathcal{R}_-^k(a, \sqrt{a(\lambda)} - \varepsilon)$ . However these estimates do not exhaust the asymptotic behaviors of the unitary group of the  $N$ -body Schrödinger operator. In fact, on the support of  $J_a$ ,  $H$  is approximated by  $H^a + T_a$ . So if the whole system has an energy  $\lambda$  and the particles in the clusters form a bound state of energy  $E^a < 0$ , the intercluster kinetic

energy is equal to  $\lambda - E^a$ . Therefore the classically forbidden region should be  $x \cdot \xi < |x| \sqrt{\lambda - E^a}$ , which is bigger than  $x \cdot \xi < |x| \sqrt{a(\lambda)}$ . To obtain this sort of precise propagation estimates is still an open problem although it would be very effective in the many body problem. However, the above estimates are sufficient to study detailed properties of the 3-body Schrödinger operator, which we explain in the next section.

## 6. Three-body problems

We shall illustrate the utility of the estimates of the resolvent or the unitary group derived in the previous sections by studying wave operators and S-matrices. In this section we restrict ourselves to the 3-body problems with short-range perturbations, namely,  $\rho$  in (2.7) is assumed to satisfy  $\rho > 1$ .

Let us first recall the following decay estimate which holds for general  $N$ -body problems with long-range perturbations:

$$\|X^{-s} e^{-itH} \varphi(H) X^{-s}\| \leq C_{s,s'} (1 + |t|)^{-s'}, \quad (6.1)$$

where  $\varphi \in C_0^\infty(\mathbf{R})$  with  $\text{supp } \varphi \subset \sigma_{\text{cont}}(H) \cap \sigma_p(H)^c \cap \Lambda^c$ ,  $0 < s' < s$  (see [39], [37]).

### 6.1. Asymptotic completeness

Taking a compact interval  $I$  in  $\sigma_{\text{cont}}(H) \cap \sigma_p(H)^c \cap \Lambda^c$ ,  $f \in C_0^\infty(I)$  and a rapidly decreasing function  $u$ , we study the asymptotic behavior of

$$u(t) = e^{-itH} f(H) u.$$

*Lemma 6.1.* *Let  $\Gamma$  be a closed cone in  $\mathcal{U}_0$  defined in (5.10). We take  $\chi \in C^\infty(\mathcal{X})$  such that  $\text{supp } \chi \subset \Gamma$  and*

$$|\partial_x^m \chi(x)| \leq C_m \langle x \rangle^{-m}, \quad \forall m \geq 0. \quad (6.2)$$

*Let  $\varphi \in C_0^\infty(\chi - \{0\})$  be such that*

$$\sup \{ \hat{x} \cdot \hat{\xi}; x \in \text{supp } \nabla \chi, |x| > 1, \xi \in \text{supp } \varphi \} < 1, \quad (6.3)$$

*where  $\hat{x} = x/|x|$ ,  $\hat{\xi} = \xi/|\xi|$ . Then if  $|I|$  is sufficiently small*

$$\left\| \frac{d}{dt} e^{itH_0} \varphi(D_x) \chi(x) u(t) \right\| = O(t^{-\rho'}) \text{ as } t \rightarrow \infty,$$

*where  $1 < \rho' < \rho$ .*

*Proof.* We compute

$$\begin{aligned} & \frac{d}{dt} e^{itH_0} \varphi(D_x) \chi(x) u(t) \\ &= i e^{itH_0} \left\{ \varphi(D_x) [H_0, \chi(x)] - \varphi(D_x) \chi(x) \sum_{(ij)} V_{ij} \right\} e^{-itH} f(H) u. \end{aligned}$$

In view of (6.1), we have

$$\left\| \chi(x) \sum_{(ij)} V_{ij} e^{-iH} f(H) u \right\| = O(t^{-\rho'}). \quad (6.4)$$

By Theorem 5.7 we have

$$\| \varphi(D_x) [H_0, \chi(x)] e^{-iH} f(H) u \| = O(t^{-m}), \quad \forall m, \quad (6.5)$$

if  $|I|$  is sufficiently small. (6.4) and (6.5) imply the lemma.  $\square$

For a small  $\varepsilon > 0$ , we take a  $C^\infty$ -function  $\rho(t)$  such that  $\rho(t) = 1$  if  $t < \varepsilon$ ,  $\rho(t) = 0$  if  $t > 2\varepsilon$ . Let  $\chi_\infty(x) \in C^\infty(\mathcal{X})$  be such that  $\chi_\infty(x) = 1$  if  $|x| > 2$ ,  $\chi_\infty(x) = 0$  if  $|x| < 1$ . For a 2-cluster decomposition  $a$ , we put

$$\chi_a(x) = \rho(|x^a|/|x|) \chi_\infty(x). \quad (6.6)$$

We also define

$$\chi_0(x) = \left( 1 - \sum_{\#a=2} \rho(|x^a|/|x|) \right) \chi_\infty(x). \quad (6.7)$$

In the following, if  $\#a = 3$ , the subscript  $a$  should be read as 0.

*Lemma 6.2.* Let  $|I|$  be sufficiently small. Then for  $\#a = 2, 3$ , there exists  $v_a \in L^2(\mathcal{X})$  such that as  $t \rightarrow \infty$ ,

$$\| \chi_a(x) u(t) - \chi_a(x) e^{-iH_a} v_a \| \rightarrow 0.$$

*Proof.* We first prove the case for  $\#a = 3$ . First we note that there exists a closed cone  $\Gamma_0 \subset \mathcal{Y}_0$  such that  $\text{supp } \chi_0 \subset \Gamma_0$ . We take a closed cone  $\Gamma_1$  such that  $\Gamma_0 \subset \Gamma_1 \subset \mathcal{Y}_0$  and a  $C^\infty$ -function  $\chi_1(x)$  homogeneous of degree 0 for  $|x| > 2$ ,  $\text{supp } \chi_1 \subset \Gamma_1$  and  $\chi_0 \chi_1 = \chi_0$ . We consider the asymptotic behavior of  $\chi_1(x) u(t)$ . Take  $g \in C_0^\infty(I)$  such that  $gf = f$ . Then in view of (6.1) and Lemma 5.1, we have

$$\begin{aligned} & \frac{d}{dt} e^{itH_0} \chi_1(x) u(t) \\ &= ie^{itH_0} [H_0, \chi_1(x)] g(H_0)^2 u(t) + O(t^{-\rho'}). \end{aligned} \quad (6.8)$$

We take  $\psi, \chi_2 \in C^\infty(\mathcal{X})$  homogeneous of degree 0 for  $|x| > \varepsilon$  such that  $\psi \nabla \chi_1 = \nabla \chi_1$ ,  $\chi_2 \psi = \psi$  and  $\text{supp } \psi, \text{supp } \chi_2$  are contained in  $\mathcal{Y}_0$ . Then by Theorem 5.7,

$$\begin{aligned} & [H_0, \chi_1(x)] g(H_0)^2 u(t) \\ &= [H_0, \chi_1(x)] g(H_0)^2 \psi(D_x) \chi_2(x) u(t) + O(t^{-1-\rho}). \end{aligned} \quad (6.9)$$

Letting  $w(t) = e^{itH_0} g(H_0) \psi(D_x) \chi_2(x) u(t)$ , we have by Lemma 6.1,

$$\left\| \frac{d}{dt} w(t) \right\| = O(t^{-\rho'}). \quad (6.10)$$



By virtue of (6.8) ~ (6.10) we have

$$\begin{aligned}
 & \frac{d}{dt} e^{itH_0} \chi_1(x) u(t) \\
 &= ie^{itH_0} [H_0, \chi_1(x)] e^{-itH_0} g(H_0) w(t) + O(t^{-\rho'}) \\
 &= \frac{d}{dt} \{ e^{itH_0} \chi_1(x) e^{-itH_0} g(H_0) w(t) \} \\
 &\quad - e^{itH_0} \chi_1(x) e^{-itH_0} g(H_0) \frac{d}{dt} w(t) + O(t^{-\rho'}) \\
 &= \frac{d}{dt} \{ e^{itH_0} \chi_1(x) e^{-itH_0} g(H_0) w(t) \} + O(t^{-\rho'}).
 \end{aligned}$$

Integrating this equation we obtain

$$e^{itH_0} \chi_1(x) u(t) = e^{itH_0} \chi_1(x) e^{-itH_0} g(H_0) w(t) + w_1(t), \quad (6.11)$$

where  $w_1(t) \rightarrow w_1 \in L^2(\mathcal{X})$  as  $t \rightarrow \infty$ . From (6.11), the assertion (1) of the present lemma follows immediately.

The proof of the case for  $\#a = 2$  is essentially the same as above if we note that  $\chi_a(x)(H - H_a) = O(\langle x \rangle^{-\rho})$  and  $\text{supp } \nabla \chi_a \subset \mathcal{Y}_0$ .  $\square$

The essential point of the above proof is that  $\text{supp } \nabla \chi_a \subset \mathcal{Y}_0$  which holds only for the 3-body problem.

We are now in a position to prove the asymptotic completeness of the wave operators. For a 2-cluster decomposition  $a$ , let  $P^a$  be the projection onto the bound states of the 2-body subsystem  $H^a = -\Delta_{x^a} + V^a(x^a)$ . Then as is well-known the wave operators

$$W_0^\pm = s - \lim_{t \rightarrow \pm \infty} e^{itH} e^{-itH_0}, \quad (6.12)$$

$$W_a^\pm = s - \lim_{t \rightarrow \pm \infty} e^{itH} e^{-itH_a} P^a \quad (6.13)$$

exist and their ranges are orthogonal.

**Theorem 6.3.** (*Asymptotic Completeness*).

$$\mathcal{H}_{ac}(H) = \sum_{\#a=2,3} \oplus \text{Ran } W_a^\pm,$$

where  $\mathcal{H}_{ac}(H)$  denotes the absolutely continuous subspace for  $H$ .

*Proof.* Let  $I$  be a small interval in  $\sigma_{\text{cont}}(H) \cap \sigma_p(H)^c \cap \Lambda^c$  and take  $f \in C_0^\infty(I)$ . Let  $u$  be a rapidly decreasing function. Then by Lemma 6.2,

$$e^{-itH} f(H) u \sim \sum_{\#a=2,3} \chi_a(x) e^{-itH_a} v_a.$$

By the completeness of 2-body wave operators we have for  $\#a = 2$

$$e^{-itH_a}v_a \sim e^{-itH_0}w_a + e^{-itH_a}P^a\tilde{w}_a.$$

We show that for any  $a, \#a = 2, 3, w \in L^2(\mathcal{X})$  and  $\varepsilon > 0$ , there exists  $u_a \in L^2(\mathcal{X})$  such that

$$\lim_{t \rightarrow \infty} \|\chi_a(x)e^{-itH_0}w - e^{-itH_0}u_a\| < \varepsilon. \quad (6.14)$$

In fact, one can take  $u$  in such a way that  $\hat{u} \in C_0^\infty(\mathcal{X} - \{0\})$  and

$$\sup_{t > 0} \|\chi_a(x)e^{-itH_0}(w - u)\| < \varepsilon/2.$$

We take  $\tilde{\chi}_a(x)$  such that  $\tilde{\chi}_a(x)$  is homogeneous of degree 0 for all  $x \neq 0$  and  $\tilde{\chi}_a(x) = \chi_a(x)$  for  $|x| > 2$ . Then

$$\lim_{t \rightarrow \infty} \|(\chi_a(x) - \tilde{\chi}_a(x))e^{-itH_0}u\| = 0.$$

Using the stationary phase method and the homogeneity of  $\tilde{\chi}_a(x)$ , we see that  $\tilde{\chi}_a(x)e^{-itH_0}u$  is asymptotically equal to

$$\text{Const. } t^{-\nu} \tilde{\chi}_a(x) \hat{u}(x/2t) = \text{Const. } t^{-\nu} \tilde{\chi}_a(x/2t) \hat{u}(x/2t).$$

By using the stationary phase method again, we see that this is asymptotically equal to  $e^{-itH_0}u_a$  for some  $u_a \in L^2(\mathcal{X})$ . This proves (6.14).

We have thus shown that

$$e^{-itH}f(H)u \sim e^{-itH_0}u_0 + \sum_{\#a=2} \chi_a(x)e^{-itH_a}P^a\tilde{w}_a.$$

We approximate  $P^a$  by a finite rank projection for the negative eigenvalues and use the exponential decay property of eigenfunctions to see that for any  $\varepsilon > 0$  there exist  $T > 0$  and  $u_a$  such that

$$\left\| e^{-itH}f(H)u - \left( e^{-itH_0}u_0 + \sum_{\#a=2} e^{-itH_a}P^a u_a \right) \right\| < \varepsilon \quad (6.15)$$

if  $t > T$ . As is well-known, (6.15) implies the theorem. In fact suppose there exist  $u \in \mathcal{H}_{ac}(H)$  orthogonal to the ranges of all channel wave operators. Take  $I$  as above. In view of (6.15), we have

$$\|f(H)u\|^2 = \left\| e^{-itH}f(H)u, e^{-itH_0}u_0 + \sum_{\#a=2} e^{-itH_a}P^a u_a \right\|^2 + O(\varepsilon).$$

Since  $u$  is orthogonal to the ranges of wave operators, we have by letting  $t$  tend to infinity,

## 6.2. Collision processes

Let us consider the collision process of three particles labelled by 1, 2, 3. It is physically natural to assume that the initial state is of 2-cluster. Then after the collision there occur the following five phenomena:

$$(12) + (3) \Rightarrow \begin{cases} (a) & (12) + (3), \\ (b) & (12)^* + (3)_*, \\ (c) & (12)' + (3), \\ (d) & (13) + (2), \\ (e) & (1) + (2) + (3). \end{cases}$$

(a) is an elastic process. In (b) the energy of the pair (12) changes, while in (c) the energies are unchanged but this pair takes a different state. (d) is a rearrangement process and (e) is a break-up process. All these processes are described by the  $S$ -matrices. There are two methods to introduce them, the time-dependent method and the stationary one. The latter is formulated as follows. Let  $a$  be a 2-cluster decomposition and  $E^a, u^a(x^a)$  be a negative eigenvalue and the associated eigenfunction of the 2-body subsystem  $H^a$ . Then  $e^{i(\lambda - E^a)^{1/2} \omega \cdot x_a} u^a(x^a)$  represents the initial state with total energy  $\lambda$  and incident direction  $\omega \in S^{v-1}$ . We seek a generalized eigenfunction  $\varphi(x, \lambda, \omega)$  such that  $(H - \lambda)\varphi = 0$  and  $v = \varphi - e^{i(\lambda - E^a)^{1/2} \omega \cdot x_a} u^a(x^a)$  satisfies the *outgoing radiation condition*. Then by observing the spatial asymptotics of  $v$  we find the  $S$ -matrices.

## 6.3. $S$ -matrices

To carry out these procedures is not an easy problem and has been dealt with mainly by physicists (see e.g. [43], [44]). In the following, we shall explain a mathematically rigorous treatment. We define the outgoing radiation condition by (4.2). (When the potentials are  $C^\infty$ , one can define the radiation condition equivalently in terms of Ps.D.Op.'s. See [33].) Then by Theorem 4.2, the generalized eigenfunction  $\varphi$  is obtained uniquely by the following formula:

$$\varphi(x, \lambda, \omega) = e^{i(\lambda - E^a)^{1/2} \omega \cdot x_a} u^a(x^a) - v, \quad (6.16)$$

$$v = R(\lambda + i0)f, \quad (6.17)$$

$$f = f(x, \lambda, \omega) = \sum_{\#b=2, b \neq a} V^b(x^b) u^a(x^a) e^{i(\lambda - E^a)^{1/2} \omega \cdot x_a}. \quad (6.18)$$

We assume that  $v = 3$ . Let

$$S_{\text{free}} = S^5 \setminus \cup_{\#b=2} X_b, \quad S_{\text{int}} = S^5 \cap (\cup_{\#b=2} X_b).$$

The suffix int means the interaction. We first show how to obtain these  $S$ -matrices from the spatial asymptotics of  $v$ . Let  $\hat{S}_{0a}(\lambda)$  be the  $S$ -matrix associated with the break-up process (e) and  $\hat{S}_{0a}(\lambda; \theta, \omega)$  be its kernel. Then we have

**Theorem 6.4** *If  $\rho > 4 + 1/2$ , for any  $\lambda > 0$ ,*

$$s - \lim_{r \rightarrow \infty} r^{5/2} e^{-i\gamma\lambda^{1/2}} v(r \cdot) = C_1(\lambda) \hat{S}_{0a}(\lambda; \cdot, \omega),$$

$$C_1(\lambda) = e^{-\pi i/4} 2\pi \lambda^{-1/4} (\lambda - E^a)^{-1/4},$$

in  $L^2_{\text{loc}}(S_{\text{free}})$ .

In the neighborhood of  $S_{\text{int}}$ , we can show the above convergence in an averaged sense. We take  $\chi_b(x) \in C^\infty(X)$  such that  $\chi_b(x) = 1$  if  $|x^b|/|x| < \varepsilon$ ,  $\chi_b(x) = 0$  if  $|x^b|/|x| > 2\varepsilon$ , and  $\rho_+(t) \in C^\infty(\mathbf{R}^1)$  such that  $\rho_+(t) = 1$  if  $t > 1 - \varepsilon$ ,  $\rho_+(t) = 0$  if  $t < 1 - 2\varepsilon$ , where  $\varepsilon$  is a small positive constant. We also take  $\rho(t) \in C^\infty_0((0, \infty))$  such that  $\int_0^\infty \rho(t) dt = 1$ .

**Theorem 6.5.** *Suppose that  $V_{ij}$ 's are rapidly decreasing functions. Let  $b$  be any 2-cluster decomposition and decompose  $\theta \in S^5$  as  $\theta = (\theta^b, \theta_b)$  in accordance with the choice of the Jacobi-coordinates. Then*

$$\begin{aligned} s - \lim_{R \rightarrow \infty} \frac{1}{R} \int_{\mathbf{R}^6} e^{-i(\lambda)^{1/2} \theta \cdot x} \theta_b \cdot \hat{x}_b \rho + \left( \frac{\theta_b}{|\theta_b|} \cdot \frac{x_b}{|x_b|} \right) \chi_b(x) \rho \left( \frac{|x_b|}{R} \right) v(x) dx \\ = C_2(\lambda) \hat{S}_{0a}(\lambda; \theta, \omega), \\ C_2(\lambda) = -(2\pi)^{7/2} \lambda^{-3/2} (\lambda - E^a)^{-1/4}, \end{aligned}$$

in  $L^2(\tilde{S}_{\text{int}}(b))$ , where  $\tilde{S}_{\text{int}}(b)$  is a small neighborhood of  $S_{\text{int}} \cap \mathcal{X}_b$  and  $\hat{x}_b = x_b/|x_b|$ .

In the neighborhood of the  $\chi_b$ -plane, there are two sorts of scattering, the 3-cluster scattering and the 2-cluster scattering. We can distinguish between them by changing the way of taking the limit at infinity of  $v$ .

Let  $u^{bl}(x^b)$  ( $1 \leq l \leq \dim P^b$ ) be a normalized eigenfunction of  $H^b$  with eigenvalue  $E_l^b \leq 0$ . Let  $A_{bl}(\lambda; \theta_b, \omega)$  be the 2-cluster scattering amplitude associated with the process in which, after the collision, the pair takes the bound state  $u^{bl}$ , which corresponds to one of the cases (a) ~ (d).

**Theorem 6.6.** *Suppose that  $V_{ij}$ 's are rapidly decreasing functions. Fix  $R > 0$  arbitrarily. Then as  $r = |x_b| \rightarrow \infty$ , we have the following asymptotic expansion*

$$\begin{aligned} v \simeq \sum_l C_{bl}(\lambda) u^{bl}(x^b) r^{-1} e^{i(\lambda - E_l^b)^{1/2} r} A_{bl}(\lambda; \theta_b, \omega), \quad \theta_b = x_b/r, \\ C_{bl}(\lambda) = 2\pi i (\lambda - E^a)^{-1/4} (\lambda - E_l^b)^{-1/4}, \end{aligned}$$

uniformly for  $|x^b| < R, \theta_b \in S^2$ .

Theorems 6.4 and 6.5 were provided in [32]. Theorem 6.6 is an improvement of Theorem 1.3 in [32].

*Proof of Theorem 6.6.* For a small  $\varepsilon > 0$ , take  $\psi_1(t), \psi_2(t) \in C^\infty(\mathbf{R})$  such that  $\psi_1(t) + \psi_2(t) = 1$ ,  $\psi_1(t) = 1$  if  $t > \lambda - \varepsilon$ ,  $\psi_1(t) = 0$  if  $t < \lambda - 2\varepsilon$ . Then in [32],  $\psi_1(|D_{x_b}|^2)v$  was shown to have the desired asymptotic expansion.

For a small  $\varepsilon' > 0$ , let  $\chi(t) \in C^\infty(\mathbf{R})$  be such that  $\chi(t) = 1$  if  $t < \varepsilon'$ ,  $\chi(t) = 0$  if  $t > 2\varepsilon'$ .

Then as  $r = |x_b| \rightarrow \infty$ ,

$$\psi_2(|D_{x_b}|^2)v = \chi(|x^b|/|x|)\psi_2(|D_{x_b}|^2)v$$

if  $|x^b| < R$ .  $P = \chi(|x^b|/|x|)\psi_2(|D_{x_b}|^2)$  is a Ps.D.Op. with symbol satisfying the assumption of Theorem 5.4. Therefore by taking  $\varepsilon'$  small enough, we see that  $P_v$  is rapidly decreasing, which completes the proof.  $\square$

As for the continuity of the kernel of  $\hat{S}_{0a}(\lambda)$ , we have the following result ([31]).

**Theorem 6.7.** (1) Suppose  $\rho > 4 + 1/2$ . Then  $\hat{S}_{0a}(\lambda)$  has a continuous kernel outside  $S_{int}$ :

$$\hat{S}_{0a}(\lambda; \theta, \omega) \in C((0, \infty) \times S_{free} \times S^2).$$

(2) Suppose  $\rho > 5 + 1/2$ . Then as  $|\theta^b| \rightarrow 0$ ,

$$\hat{S}_{0a}(\lambda; \theta, \omega) \simeq |\theta^b|^{-1} A_{b,-1}\left(\lambda; \frac{\theta^b}{|\theta^b|}, \theta_b, \omega\right) + A_{b,0}\left(\lambda; \frac{\theta^b}{|\theta^b|}, \theta_b, \omega\right),$$

where

$$\begin{aligned} & A_{b,-1}\left(\lambda; \frac{\theta^b}{|\theta^b|}, \theta_b, \omega\right) \\ &= \sum_j^{\text{finite}} C_{b1}^{(j)}(\lambda; \theta_b, \omega) \times \int_{\mathbf{R}^3} \frac{\theta^b}{|\theta^b|} \cdot x^b V^b(x^b) u_j^b(x^b) dx^b \\ &+ C_{b2}(\lambda; \theta_b, \omega) \times \int_{\mathbf{R}^3} V^b(x^b) \varphi^b(x^b) dx^b, \end{aligned}$$

$u_j^b$  being the eigenfunction with zero eigenvalue for  $H^b$ , and  $\varphi^b$  the zero-resonance.  $A_{b,0}$  is continuous with respect to all of its arguments.  $A_{b,-1} = 0$ , if 0 is neither an eigenvalue nor the resonance for  $H^b$ . In this case,  $\hat{S}_{0a}(\lambda; \theta, \omega)$  is continuous at  $\theta^b = 0$ .

(3) Up to a multiplicative constant depending only on  $\lambda$  and  $E^a$ ,  $C_{b1}^{(j)}(\lambda; \theta_b, \omega)$  and  $C_{b2}(\lambda; \theta_b, \omega)$  coincide with the scattering amplitudes for two cluster scattering.

In the proof of all these theorems, Theorem 5.7 plays a crucial role.

We end this section by noting that concerning the collision process from 3-clusters to 3-clusters:

$$(1) + (2) + (3) \Rightarrow (1) + (2) + (3),$$

we do not know even the existence of the scattering kernel.

## 7. S-matrices in $N$ -body problems

### 7.1. Scattering from 2-clusters to $N$ -clusters

The results in §6 can be generalized to the  $N$ -body problem. Let us consider the break-up process from 2-clusters to  $N$ -clusters:

$$(1 \dots k) + (k+1 \dots N) \Rightarrow (1) + (2) + \dots + (N).$$

We formulate the corresponding  $S$ -matrix by the time-dependent method. Let  $a$  be a 2-cluster decomposition and  $E^a, u^a(x^a)$  be a negative eigenvalue and the associated eigenfunction of  $H^a$ . Let  $W_0^\pm$  and  $W_a^\pm$  be the wave operators defined by

$$W_0^\pm = s - \lim_{t \rightarrow \pm \infty} e^{itH} e^{-itH_0}, \quad (7.1)$$

$$W_a^\pm = s - \lim_{t \rightarrow \pm \infty} e^{itH} e^{-itH_a} J_a, \quad (7.2)$$

where  $J_a: L^2(\mathcal{X}_a) \rightarrow L^2(\mathcal{X})$  is the injection defined by  $(J_a f)(x^a, x_a) = u^a(x^a) f(x_a)$ . Let  $S_{0a}$  be the scattering operator defined by

$$S_{0a} = (W_0^+)^* W_a^-. \quad (7.3)$$

Introducing the Fourier transformations  $\mathcal{F}_a: L^2(\mathcal{X}_a) \rightarrow L^2((E^a, \infty); L^2(S^{v-1}))$  and  $\mathcal{F}_0: L^2(\mathcal{X}) \rightarrow L^2((0, \infty); L^2(S^{n-1}))$ ,  $n = \dim \mathcal{X}$ , by

$$(\mathcal{F}_a f)(\lambda, \omega) = 2^{-1/2} (2\pi)^{-v/2} (\lambda - E^a)^{(v-2)/4} \int_{\mathcal{X}_a} e^{-i(\lambda - E^a)^{1/2} \omega \cdot x_a} f(x_a) dx_a,$$

$$(\mathcal{F}_0 f)(\lambda, \theta) = 2^{-1/2} (2\pi)^{-n/2} \lambda^{(n-2)/4} \int_{\mathcal{X}} e^{-i(\lambda)^{1/2} \theta \cdot x} f(x) dx,$$

we consider the Fourier transform of  $S_{0a}$ :

$$\hat{S}_{0a} = \mathcal{F}_0 S_{0a} \mathcal{F}_a^*. \quad (7.4)$$

Then  $\hat{S}_{0a}$  is represented by the direct integral ([46])

$$\hat{S}_{0a} = \int_0^\infty \oplus \hat{S}_{0a}(\lambda) d\lambda, \quad \hat{S}_{0a}(\lambda) \in \mathbf{B}(L^2(S^{v-1}); L^2(S^{n-1})).$$

This  $\hat{S}_{0a}(\lambda)$  is called the  $S$ -matrix. Note that by this definition,  $\hat{S}_{0a}(\lambda)$  is defined for a.e.  $\lambda > 0$ .

## 7.2. Smoothness of the scattering kernel.

Let

$$S_{\text{free}} = S^{n-1} \setminus \bigcup_{2 \leq \#b \leq N-1} \mathcal{X}_b, \quad (7.5)$$

$$S_{\text{int}} = S^{n-1} \cap \left( \bigcup_{2 \leq \#b \leq N-1} \mathcal{X}_b \right), \quad (7.6)$$

and consider the regularity of the kernel of  $\hat{S}_{0a}(\lambda)$ . For a small  $\varepsilon > 0$  we define

$$S_{\text{free}}^\varepsilon = \{x \in S_{\text{free}}; |x^i - x^j| > \varepsilon \forall i \neq j\}$$

and take  $\psi(\theta) \in C^\infty(S^{n-1})$  such that

$$\psi(\theta) = \begin{cases} 1 & \text{if } \theta \in S_{\text{free}}^{4\varepsilon} \\ 0 & \text{if } \theta \notin S_{\text{free}}^{3\varepsilon} \end{cases}$$

We take  $\chi(x) \in C^\infty(\mathcal{X})$  such that for  $|x| > 1$ ,

$$\chi(x) = \begin{cases} 1 & \text{if } x/|x| \in S_{\text{free}}^{2\varepsilon} \\ 0 & \text{if } x/|x| \notin S_{\text{free}}^e \end{cases}$$

Fix  $\lambda > 0$  and take  $\varphi(t) \in C_0^\infty((0, \infty))$  such that  $\varphi(t) = 1$  near  $t = \lambda$ . Let  $P$  be a Ps.D.Op. with symbol

$$p(x, \xi) = \chi(x) \psi(\xi/|\xi|) \varphi(|\xi|^2),$$

and put

$$G = HP - PH_0 = [H_0, P] + \sum_{(ij)} V_{ij} P,$$

$$Q_a = (H - H_a) J_a.$$

Then by [31] (3.10), the following formula holds:

$$\begin{aligned} \psi(\theta) \hat{S}_{0a}(\lambda) &= -2\pi i \mathcal{F}_0(\lambda) P^* Q_a \mathcal{F}_a^*(\lambda) \\ &\quad + 2\pi i \mathcal{F}_0(\lambda) G^* R(\lambda + i0) Q_a \mathcal{F}_a^*(\lambda). \end{aligned}$$

Here we note that  $[H_0, P]$  is a Ps.D.Op. satisfying the assumption in Theorem 5.7. Therefore if  $V_{ij}$  is rapidly decreasing  $G^* R(\lambda + i0) Q_a$  maps a polynomially increasing function to a rapidly decreasing one. This implies that  $\psi(\theta) \hat{S}_{0a}(\lambda)$  has a  $C^\infty$ -kernel when  $V_{ij}$  is rapidly decreasing. This fact was extended to general smooth short-range potentials by Skibsted [51].

**Theorem 7.1.** *Suppose that  $V_{ij}$ 's satisfy the assumption in 4.3 in §4. Then  $\hat{S}_{0a}(\lambda)$  has a  $C^\infty$ -kernel outside  $S_{\text{int}}$ :*

$$\hat{S}_{0a}(\lambda; \theta, \omega) \in C^\infty((0, \infty) \times S_{\text{free}} \times S^{v-1}).$$

In [8], Bommier proved the smoothness of the kernel of the scattering amplitude associated with the process from 2-clusters to 2-clusters for long-range potentials.

### 7.3. Singularities of the scattering kernel

In view of the results for the 3-body problem, one may well to conjecture that  $\hat{S}_{0a}(\lambda; \theta, \omega)$  has singularities on the variety  $S_{\text{int}}$ . At the present stage, it is not easy to determine completely these singularities because of the lack of the precise resolvent estimates for the  $N$ -body problem and also of informations for the zero-eigenvalue and the zero-resonances of the subsystems. We can investigate them only on some subvariety of  $S_{\text{int}}$ . For a cluster decomposition  $b$  we define

$$S_{\text{int}}^*(b) = \{x \in S_{\text{int}}; x^i \neq x^j \text{ if } (ij) \notin b\}.$$

We assume that  $v = 3$ , hence  $n = 3(N - 1)$ . When  $\#b = N - 1$ , the behavior of  $\hat{S}_{0a}(\lambda; \theta, \omega)$  near  $S_{\text{int}}^*(b)$  is the same as in the case of the 3-body problem.

**Theorem 7.2.** (1) *Let  $\#b = N - 1$ . Suppose  $\rho > (n + 5)/2$ . Then around  $S_{\text{int}}^*(b)$  the*

following asymptotic expansion holds:

$$\hat{S}_{0a}(\lambda; \theta, \omega) \simeq |\theta^b|^{-1} A_{b,-1} \left( \lambda; \frac{\theta^b}{|\theta^b|}, \theta_b, \omega \right) + A_{b,0} \left( \lambda; \frac{\theta^b}{|\theta^b|}, \theta_b, \omega \right),$$

where

$$\begin{aligned} & A_{b,-1} \left( \lambda; \frac{\theta^b}{|\theta^b|}, \theta_b, \omega \right) \\ &= \sum_j^{\text{finite}} C_{b1}^{(j)}(\lambda; \theta_b, \omega) \times \int_{\mathbf{R}^3} \frac{\theta^b}{|\theta^b|} \cdot x^b V^b(x^b) u_j^b(x^b) dx^b \\ &+ C_{b2}(\lambda; \theta_b, \omega) \times \int_{\mathbf{R}^3} V^b(x^b) \varphi^b(x^b) dx^b, \end{aligned}$$

$u_j^b$  being the eigenfunction with zero eigenvalue for  $H^b$ , and  $\varphi^b$  the zero-resonance.  $A_{b,0}$  is continuous with respect to all of its arguments.  $A_{b,-1} = 0$ , if 0 is neither an eigenvalue nor the resonance for  $H^b$ . In this case,  $\hat{S}_{0a}(\lambda; \theta, \omega)$  is continuous on  $S_{\text{int}}(b)$ .

(2) Up to a multiplicative constant depending only on  $\lambda$  and  $E^a$ ,  $C_{b1}^{(j)}(\lambda; \theta_b, \omega)$  and  $C_{b2}(\lambda; \theta_b, \omega)$  coincide with the scattering amplitudes for two cluster scattering.

We explain the outline of the proof of Theorem 7.2 which is essentially the same as that in §4 of [31].

We fix  $\lambda > 0$  and take small constants  $\varepsilon_1, \varepsilon_2 > 0$ , open sets  $\mathcal{O}^b \subset \mathcal{X}^b$ ,  $\mathcal{O}_b \subset \mathcal{X}_b$  having the following properties;

$$\mathcal{O} \equiv \mathcal{O}^b \times \mathcal{O}_b \subset \{\xi; \lambda - \varepsilon_1 < |\xi|^2 < \lambda + \varepsilon_1\},$$

$$\bar{\mathcal{O}}^b \subset \{\xi^b; |\xi^b| < \varepsilon_2\},$$

$$\bar{\mathcal{O}}_b \subset \mathcal{X}_b \setminus \bigcup_{c \neq b} \mathcal{X}_c.$$

Let  $\varphi(\xi)$  be such that

$$\varphi(\xi) = \varphi^b(\xi^b) \varphi_b(\xi_b),$$

where  $\varphi^b \in C_0^\infty(\mathcal{O}^b)$ ,  $\varphi_b \in C_0^\infty(\mathcal{O}_b)$ . For a subset  $A$  in  $\mathbf{R}^k$  we define a conic subset  $\Gamma(A)$  in  $\mathbf{R}^k$  by

$$\Gamma(A) = \{tx; t > 0, x \in A\}.$$

Let  $\tilde{\mathcal{O}}$  be a small neighborhood of  $\mathcal{O}$ . We take  $\chi(x) \in C^\infty(\mathcal{X})$  homogeneous of degree 0 for  $|x| > 1$  satisfying  $\chi(x) = 1$  if  $|x| > 1$  and  $x \in \Gamma(\mathcal{O})$ ,  $\chi(x) = 0$  if  $x \notin \Gamma(\tilde{\mathcal{O}})$ . Then by the stationary phase method we have

$$W_0^+ \varphi(D_x) = s - \lim_{t \rightarrow \infty} e^{itH} \chi(x) \varphi_b(D_{x_b}) e^{-itH_0} \varphi^b(D_{x^b}).$$

We define

$$P_b = \chi(x) \varphi_b(D_{x_b}),$$

$$G = HP_b - P_b H_0.$$

Then by the same arguments as in the proof of [31] Lemma 3.1, one can prove the



following formula of the localisation of the S-matrix: Let  $\hat{f} \in C_0^\infty((E^a, \infty); C^\infty(S^2))$  and  $\hat{g} \in C_0^\infty((0, \infty); C^\infty(S^{n-1}))$ . Then

$$\begin{aligned} & (\mathcal{F}_0 \varphi(D_x) S_{0a} \mathcal{F}_a^* \hat{f}, \hat{g}) \\ &= -2\pi i \int_0^\infty \langle \mathcal{F}_0(\lambda) \varphi^b(D_{x_b}) P_b^* Q_a \mathcal{F}_a^*(\lambda) \hat{f}(\lambda), \hat{g}(\lambda) \rangle d\lambda \\ &+ \lim_{\varepsilon \downarrow 0} 2\pi i \int_0^\infty \langle \mathcal{F}_0(\lambda) \varphi^b(D_{x_b}) G^* R(\lambda + i\varepsilon) Q_a \mathcal{F}_a^*(\lambda) \hat{f}(\lambda), \hat{g}(\lambda) \rangle d\lambda, \quad (7.7) \end{aligned}$$

where  $(\mathcal{F}_0(\lambda)f)(\theta) = (\mathcal{F}_0 f)(\lambda, \theta)$ ,  $(\mathcal{F}_a(\lambda)f)(\omega) = (\mathcal{F}_a f)(\lambda, \omega)$  and  $\langle \cdot, \cdot \rangle$  is the inner product of  $L^2(S^{n-1})$ . Since the first term of the right-hand side of (7.7) is easily seen to have a continuous kernel, we devote ourselves to study the second term.

Let  $I_b = H - H_b$ . Then  $G$  consists of three terms

$$G = I_b P_b + [H_0, P_b] + V^b P_b.$$

Since on the support of  $\chi(x)$ ,  $I_b(x)$  behaves like  $\langle x \rangle^{-\rho}$ , the term  $I_b P_b$  gives rise to a continuous kernel.

To show that  $[H_0, P_b]$  gives rise to a continuous kernel we have only to prove

*Lemma 7.3.* For  $1/2 < s' < s$ ,

$$\sup_{\varepsilon > 0} \|\langle x \rangle^{s'} [H_0, P_b] R(\lambda + i\varepsilon) \langle x \rangle^{-s}\| < \infty.$$

*Proof.* Let  $\tilde{\mathcal{O}}^b$  and  $\tilde{\mathcal{O}}_b$  be small neighborhoods of  $\mathcal{O}^b$  and  $\mathcal{O}_b$  respectively. We take  $\chi(x)$  in such a way that for  $|x| > 1$ ,

$$\chi(x) = \chi^b(x^b/|x|) \chi_b(x_b/|x|),$$

where  $\chi^b(y^b) = 1$  if  $y^b \in \Gamma(\mathcal{O}^b)$ ,  $\chi^b(y^b) = 0$  if  $y^b \notin \Gamma(\tilde{\mathcal{O}}^b)$ ,  $\chi_b(y_b) = 1$  if  $y_b \in \Gamma(\mathcal{O}_b)$ ,  $\chi_b(y_b) = 0$  if  $y_b \notin \Gamma(\tilde{\mathcal{O}}_b)$ . Then for  $|x| > 1$

$$\begin{aligned} [H_0, P_b] &= [H_0, \chi^b(x^b/|x|)] \chi_b(x_b/|x|) \varphi_b(D_{x_b}) \\ &+ \chi^b(x^b/|x|) [H_0, \chi_b(x_b/|x|)] \varphi_b(D_{x_b}). \end{aligned}$$

On the support of the symbol of  $[H_0, \chi_b(x_b/|x|)] \varphi_b(D_{x_b})$ ,  $x_b \cdot \xi_b < (1 - \varepsilon) \langle x_b \rangle |\xi_b|$  for some  $\varepsilon > 0$  and  $|\xi_b|$  is very close to  $\sqrt{\lambda}$ . Therefore one can apply Theorem 5.4 to this operator.

On the support of the symbol of  $[H_0, \chi^b(x^b/|x|)] \chi_b(x_b/|x|) \varphi_b(D_{x_b})$ ,  $x$  belongs to  $\mathcal{U}_0$ . We take  $\psi(t) \in C_0^\infty((0, \infty))$  such that  $\psi(t) = 1$  if  $|t - \lambda| < \delta$ ,  $\psi(t) = 0$  if  $|t - \lambda| > 2\delta$ . Then by taking  $\delta$  sufficiently small and using Lemma 5.1, one can see that

$$[H_0, \chi^b(x^b/|x|)] \chi_b(x_b/|x|) \varphi_b(D_{x_b}) \psi(H) = X^{-1} \tilde{P} + P_N,$$

where  $\tilde{P}$  is a Ps.D.Op. satisfying the assumption in Corollary 5.8 and  $P_N \in \mathcal{OP}^{-N}(X)$ ,  $N$  being sufficiently large.

The present lemma then follows from Theorem 5.4 and Corollary 5.8.  $\square$

It remains to consider the contribution of the term  $V^b P_b$ . Let

$$u = P_b R(\lambda + i\varepsilon) f, \quad (7.8)$$

$$f = I_a(x) u^a(x^a) e^{i(\lambda - E^a)^{1/2} \omega \cdot x_a}. \quad (7.9)$$

**Lemma 7.4.** *Let  $R_b(z) = (H_b - z)^{-1}$ . Then*

$$u = R_b(\lambda + i\varepsilon) g_\varepsilon,$$

where  $g_\varepsilon \in L^{2,s}$  for any  $s < \rho - 3/2$  uniformly in  $\varepsilon > 0$ .

*Proof.* A direct calculation shows that

$$\begin{aligned} g_\varepsilon &= (H_b - (\lambda + i\varepsilon)) P_b R(\lambda + i\varepsilon) f \\ &= P_b f - P_b I_b R(\lambda + i\varepsilon) f + [H_0, P_b] R(\lambda + i\varepsilon) f. \end{aligned}$$

The lemma then follows from Lemma 7.3 and the previous arguments.  $\square$

It follows from Lemma 7.4 that

$$\mathcal{F}_0(\lambda) \phi^b(D_{x^b}) V^b P_b R(\lambda + i\varepsilon) f = \mathcal{F}_0(\lambda) \phi^b(D_{x^b}) V^b R_b(\lambda + i\varepsilon) g_\varepsilon.$$

Here we recall that  $\phi^b(\theta^b) = 1$  if  $|\theta^b|$  is sufficiently small and that  $\mathcal{F}_0(\lambda)$  is the Fourier transformation in  $x$ . So if we integrate in  $x_b$ , this gives a partial Fourier transform and hence letting  $R^b(z) = (H^b - z)^{-1}$ , we have for small  $|\theta^b|$

$$\begin{aligned} &\mathcal{F}_0(\lambda) \phi^b(D_{x^b}) V^b P_b R(\lambda + i\varepsilon) f \\ &= C(\lambda) \int_{\mathcal{R}^b} e^{-i(\lambda)^{1/2} \theta^b \cdot x^b} V^b R^b(\lambda |\theta^b|^2 + i\varepsilon) \hat{g}_\varepsilon dx^b, \end{aligned} \quad (7.10)$$

$$\hat{g}_\varepsilon = \hat{g}_\varepsilon(x^b; \lambda, \theta_b, \omega) = \int_{\mathcal{R}^b} e^{-i(\lambda)^{1/2} \theta_b \cdot x_b} \hat{g}_\varepsilon(x^b, x_b) dx_b.$$

If  $\rho > (n+5)/2$ ,  $\hat{g}_\varepsilon(\cdot; \lambda, \theta_b, \omega)$  is an  $L^{2,s}(\mathbf{R}^3)$ -valued continuous function of some  $s > 5/2$ .

The formula (7.10) clarifies the reason why the zero-eigenvalue and the zero-resonance of the 2-body subsystem  $H^b$  come in as singularities of the scattering kernel. We use the following result due to Jensen-Kato [38].

**Lemma 7.5.** *If  $\rho > 5$  and  $s > 5/2$ , we have the following asymptotic expansion in  $\mathbf{B}(L^{2,s}; L^{2,-s})$ :*

$$R^b(z) = \frac{B_{-2}}{z} + \frac{B_{-1}}{\sqrt{z}} + O(1) \text{ as } z \rightarrow 0,$$

where  $B_{-2} = -P_0$ ,  $B_{-1} = -iP_0 V^b K V^b P_0 + i\langle \cdot, \varphi \rangle \varphi$ ,  $P_0$  being the projection onto the eigenspace of  $H^b$  with zero-eigenvalue,  $K$  being an integral operator with kernel  $|x^b - y^b|^2/(24\pi)$  and  $\varphi$  being the zero-resonance.

The low energy asymptotics of the resolvent depends largely on the space dimension. This is the reason we restricted ourselves to the case  $\nu = 3$ . Now the rest of the proof of Theorem 7.2 is the same as in [31], so the details are omitted.

#### 7.4. Analytic continuation of $S$ -matrices

When the pair potentials are (dilation) analytic, the  $S$ -matrices are expected to be continued meromorphically into the complex region. For the 3-body problem this was proved by Balslev [5], [6] by using the Faddeev equation. Based on the method of parametrices developed for the 2-body problem, Bommier has recently proved that the above mentioned  $S$ -matrix  $\hat{S}_{0a}(\lambda)$  has an analytic continuation in the complex region ([9]). He further studied the same problem for other  $S$ -matrices restricted to some energy regions.

#### 7.5. Total cross-sections

The channel is defined as a triple  $\alpha = \{a, E_j^a, P_j^a\}$ , where  $a$  is a cluster decomposition,  $E_j^a$  and  $P_j^a$  are an eigenvalue and the associated 1-dimensional eigenprojection of  $H^a$ . When  $\#a = N$ , we define  $\alpha = \{a, 0, 1\}$  for the sake of convenience. Let  $\hat{S}_{\beta\alpha}(\lambda)$  be the  $S$ -matrix associated with the collision process of initial channel  $\alpha$  and final channel  $\beta$ . We restrict ourselves to the initial state of 2-clusters. Then the total cross-section is defined by

$$\sigma_a(\lambda) = \sum_{\beta} \|\hat{S}_{\beta\alpha}(\lambda) - \delta_{\beta\alpha}\|_{\text{HS}}^2,$$

where  $\|\cdot\|_{\text{HS}}$  denotes the Hilbert-Schmidt norm and the summation ranges over all channels. Amrein-Pearson-Sinha [3] proved that  $\sigma_a(\lambda) < \infty$  for a.e.  $\lambda$  under the assumption that each two-body potential decays faster than  $|x|^{-(v+1)/2}$ . It is natural to expect that  $\sigma_a(\lambda)$  is finite for all energy  $\lambda$ , but this is proved only for some restricted cases ([4], [31]). This is because it is very difficult to prove the existence of the scattering kernel continuous with respect to energy for all collision processes.

One can also consider the total cross-section with fixed incident direction  $\omega$  defined by

$$\sigma_a(\lambda, \omega) = \sum_{\beta} \int |(\hat{S}_{\beta\alpha}(\lambda) - \delta_{\beta\alpha})(\theta_b, \omega)|^2 d\theta_b.$$

By the same reasoning as above it is not easy to investigate this function pointwise. Ito and Tamura [34], [35] studied the asymptotic behavior of  $\sigma_a(\lambda, \omega; h)$  as a distribution with respect to  $\lambda$  and  $\omega$  when Planck's constant  $h$  tends to 0.

An alternative approach had been introduced by Enss-Simon [16]. Their idea is based on the following observation which is valid for the 2-body problem. For any  $g \in C_0^\infty((E^a, \infty); \mathbb{C})$ , let  $g_\omega(x_a) = \hat{g}(x_a \cdot \omega)$ ,  $\hat{g}$  being the Fourier transform of  $g$ . Then  $\sigma_a(\lambda, \omega)$  can be defined through the relation

$$\int \sigma_a(\lambda, \omega) |g(\lambda)|^2 d\lambda = \sum_{\beta} \|(S_{\beta\alpha} - \delta_{\beta\alpha})g_\omega\|^2.$$

Adopting this approach Robert and Wang [47] derived the pointwise asymptotics of  $\sigma_a(\lambda, \omega; h)$  under the non-trapping condition on the classical Hamiltonian. We must also mention the recent work of Wang [56] in which is studied the high-energy asymptotics of the total cross-section using the micro-local estimates of the resolvent.

## Acknowledgement

The author expresses his sincere gratitude to Professors K B Sinha, K Maddaly and R Ramachandran for their hospitality during his stay in India.

## References

- [1] Agmon S, Spectral properties of Schrödinger operators and scattering theory, *Ann. Scuola. Norm. Sup. Pisa. Ser. 4* (1975) 151–218
- [2] Agmon S, Some new results in spectral and scattering theory of differential operators on  $L^2(\mathbb{R}^n)$ , *Séminaire Goulaouic-Schwartz*, Ecole Polytechnique, 1978–1979
- [3] Amrein W O, Pearson D B and Sinha K B, Bounds on the total scattering cross section for N-body systems, *Nuovo Cimento. 52 A* (1979) 115–131
- [4] Amrein W O and Sinha K B, On the three body scattering cross sections, *J. Phys. A: 15* (1982) 1567–1586
- [5] Balslev E, Analytic scattering theory of quantum mechanical three-body systems, *Ann. Inst. Henri Poincaré, A 32* (1980) 125–160
- [6] Balslev E, Analytic scattering theory for many body systems below the smallest three-body thresholds, *Commun. Math. Phys., 77* (1980) 173–210
- [7] Beals R, Characterization of pseudodifferential operators and applications, *Duke Math. J., 44* (1977) 45–57
- [8] Bommier A, Propriétés de la matrice de diffusion, 2-amas-2-amas, pour les problèmes à N-corps à longue portée, *Ann. Inst. Henri Poincaré Phys. Theory. 59* (1993) 237–267
- [9] Bommier A, Régularité et prolongement méromorphe de la matrice de diffusion pour les problèmes à N-corps à longue portée, Thèse de doctrat, Centre de Math., Ecole Polytechnique (1993).
- [10] Dereziński J, A new proof of the propagation theorem for N-body quantum systems, *Commun. Math. Phys., 122* (1989) 203–231
- [11] Dereziński J, Algebraic approach to the N-body long-range scattering, *Rev. Math. Phys., 3* (1991) 1–62
- [12] Dereziński J, Asymptotic completeness for N-particle long-range quantum systems, *Ann. Math., 138* (1993) 427–476
- [13] Eidus D M, The principle of limit amplitude, *Russian Math. Surv., 24* (1969) 97–167
- [14] Enss V, Completeness of three-body quantum scattering, in *Dynamics and Processes*, (eds.) P Blanchard and L Streit, Lecture Notes in Math. 1031 (1983), pp 62–83, (Berlin-Heidelberg-New York: Springer)
- [15] Enss V, Long range scattering of two-and three-body systems, *Proc. conf. Equations aux dérivées partielles, Saint Jean de Monts*, Centre de Math., Ecole Polytechniques (1989)
- [16] Enss V and Simon B, Finite total cross sections in non-relativistic quantum mechanics, *Commun. Math. Phys., 76* (1980) 177–209
- [17] Froese R G and Herbst I, Exponential bounds and absence of positive eigenvalues of N-body Schrödinger operators, *Commun. Math. Phys., 87* (1982) 429–447
- [18] Froese R G and Herbst I, A new proof of the Mourre estimate, *Duke Math. J., 49* (1982) 1075–1085
- [19] Gérard C, Sharp propagation estimates for N-particle systems, *Duke Math. J., 67* (1992) 483–515
- [20] Gérard C, Asymptotic completeness for 3-particle long-range systems, *Invent. Math., 114* (1993) 333–397
- [21] Gérard C, Distortion analyticity for N-particle Hamiltonians, *Helvetica Phys. Acta, 66* (1993) 216–225
- [22] Gérard C, Isozaki H and Skibsted E, Commutator algebra and resolvent estimates, in *Adv. Studies in Pure Mathematics 23* (1994) *Spectral and Scattering Theory and Related Topics*, (ed.) K Yajima
- [23] Gérard C, Isozaki H and Skibsted E, N-body resolvent estimates (preprint) (1993)
- [24] Gérard C and Sigal I M, Space time picture of semi classical resonances, *Commun. Math. Phys., 145* (1992) 281–328
- [25] Graf G M, Asmptotic completeness for N-body short range systems: a new proof, *Commun. Math. Phys., 132* (1990) 73–101
- [26] Grushin V V, On Sommerfeld type conditions for a certain class of partial differential equations, *AMS Transl. Ser. 2. 51* (1966) 82–112

- [27] Helffer B and Sjöstrand J, Equation de Schrödinger avec champ magnétique et équation de Harper, *Lecture Notes in Physics*, 345, *Schrödinger Operators*, H. Holden A. Jensen (eds) (Berlin-Heidelberg-New York Springer) (1989) pp. 118–197
- [28] Hörmander L, *The analysis of linear partial differential operators* Vol 4 (Berlin-Heidelberg-New York Springer)
- [29] Ikebe T and Saito Y, Limiting absorption method and absolute continuity for the Schrödinger operator, *J. Math. Kyoto Univ.* **7** (1972) 513–542
- [30] Isozaki H, Differentiability of generalized Fourier transforms associated with Schrödinger operators, *J. Math. Kyoto Univ.* **25** (1985) 789–806
- [31] Isozaki H, Structures of S-matrices for three-body Schrödinger operators, *Commun. Math. Phys.* **146** (1992) 241–258
- [32] Isozaki H, Asymptotic properties of generalized eigenfunctions for three body Schrödinger operators, *Commun. Math. Phys.*, **153** (1993) 1–21
- [33] Isozaki H, A generalization of the radiation condition of Sommerfeld for  $N$ -body Schrödinger operators *Duke Math. J.* **74** (1994) 557–584
- [34] Ito H T and Tamura H, Semi-classical asymptotics for total scattering cross sections of 3-body systems, *J. Math. Kyoto Univ.*, **32** (1992) 533–555
- [35] Ito H T and Tamura H, Semi-classical asymptotics for total scattering cross-sections of  $N$ -body quantum systems (preprint) (1992)
- [36] Jäger W, Ein gewöhnlicher Differential-operator zweiter Ordnung für Funktionen mit Werte in einem Hilbertraum, *Math. Z.* **113** (1970) 68–98
- [37] Jensen A, Propagation estimates for Schrödinger-type operators, *Trans. Am. Math. Soc.*, **291** (1985) 129–144
- [38] Jensen A and Kato T, Spectral properties of Schrödinger operators and time decay of the wave functions, *Duke Math. J.* **46** (1979) 583–611
- [39] Jensen A, Mourre E and Perry P, Multiple commutator estimates and resolvent smoothness in quantum scattering theory, *Ann. Inst. Henri Poincaré, Physique Théorique*, **41** (1984) 207–225
- [40] Kuroda S T, Scattering theory for differential operators, I and II, *J. Math. Soc. Jpn.*, **25** (1972) 75–104 and 222–234
- [41] Mourre E, Absence of singular continuous spectrum of certain self-adjoint operators, *Commun. Math. Phys.*, **78** (1981) 391–408
- [42] Mourre E, Opérateurs conjugués et propriétés de propagations, *Commun. Math. Phys.*, **91** (1983) 279–300
- [43] Newton R G, The asymptotic form of three-particle wave functions and the cross sections, *Ann. Phys.*, **74** (1972) 324–351
- [44] Nuttall J, Asymptotic form of the three-particle scattering wave functions for free incident particles, *J. Math. Phys.* **12** (1971) 1896–1899
- [45] Perry P, Sigal I M and Simon B, Spectral analysis of  $N$ -body Schrödinger operators, *Ann. Math.*, **144** (1981) 519–567
- [46] Reed M and Simon B, *Methods of Modern Mathematical Physics*, 4, (New York-San Francisco-London Academic Press) (1979)
- [47] Robert D and Wang X P, Pointwise semiclassical asymptotics for total cross-sections in  $N$ -body problems (preprint) Université de Nantes (1992)
- [48] Sigal I M and Soffer A,  $N$ -particle scattering problem: Asymptotic completeness for short range systems, *Ann. Math.*, **125** (1987) 35–108
- [49] Sigal I M and Soffer A, Local decay and propagation estimates for time dependent and time independent Hamiltonians (preprint) Princeton University (1988)
- [50] Skibsted E, Propagation estimates for  $N$ -body Schrödinger operators, *Commun. Math. Phys.*, **142** (1991) 67–98
- [51] Skibsted E, Smoothness of  $N$ -body scattering amplitudes, *Rev. Math. Phys.*, **4** (1992) 619–658
- [52] Soffer A, On the many body problem in quantum mechanics, *S.M.F. Astérisque*, **207** (1992) 109–152
- [53] Tamura H, Asymptotic completeness for  $N$ -body Schrödinger operators with shortrange interactions, *Comm. P.D.E.*, **16** (1991) 1129–1154
- [54] Wang X P, On the three-body long-range scattering problems, *Lett. Math. Phys.*, **25** (1992) 267–276
- [55] Wang X P, Micro-local resolvent estimates for  $N$ -body Schrödinger operators, *J. Fac. Sci. Univ. Tokyo, Sect. 1A, Math.*, **40** (1993) 337–385
- [56] Wang X P, Total cross sections in  $N$ -body problems: Finiteness and high energy asymptotics, *Commun. Math. Phys.* **156** (1993) 333–354



# A conjecture for some partial differential operators on $L^2(R^n)$

PL. MUTHURAMALINGAM

Indian Statistical Institute, R.V. College Post, Bangalore 560059, India

**Abstract.** On  $\mathcal{X} = L^2(R^n)$ , let  $Q = (Q_1, Q_2, \dots, Q_n)$  and  $P = (P_1, P_2, \dots, P_n)$  be the operators given by  $(Q_j f)(x) = x_j f(x)$ ,  $P_j = -i\partial/\partial x_j$ . For any  $C^\infty$  function  $h: R^n \rightarrow R$  put  $H_0 = h(P)$  and  $H = H_0 + (1 + Q^2)^{-\delta}$ , where  $\delta > 1/2$ . By the method of scattering theory we prove that  $H_{ac}$ , the absolutely continuous part of  $H$  is unitarily equivalent to  $H_0$  when (a)  $n = 1$  and (b) for  $n \geq 2$ , when  $h$  is in a large class of polynomials. It is conjectured that the results are true for any polynomial  $h$ . We use the techniques of Enss' method and the idea of bound states for momentum.

**Keywords.** Partial differential operators; scattering theory; Enss' method.

## 1. The conjecture and justification for the conjecture

Let  $\mathcal{X} = L^2(R^n)$  be the Hilbert space of all square integrable functions on  $R^n$ . Let  $Q = (Q_1, Q_2, \dots, Q_n)$  and  $P = (P_1, P_2, \dots, P_n)$  be the position and momentum operators on  $L^2(R^n)$  given by  $(Q_j f)(x) = x_j f(x)$ , and  $(P_j f)(x) = -i(D_j f)(x)$ ,  $D_j = \frac{\partial}{\partial x_j}$ . Let  $h: R^n \rightarrow R$  be any  $C^\infty$  function such that  $h$  and all its derivatives have at most polynomial growth. Put  $H_0 = h(P)$ . Let  $W(x) = \langle x \rangle^{-1-\delta}$  for some  $\delta > 0$  where  $\langle x \rangle = (1 + x^2)^{1/2}$ . Let  $H = H_0 + W(Q)$ . Let  $U_t$  and  $V_t$  be the free and total unitary evolution groups given by the self-adjoint operators  $H_0$  and  $H$  viz

$$U_t = \exp[-it H_0], \quad V_t = \exp[-it H]$$

*Conjecture.* Let  $h$  be any polynomial with  $\sum_{\alpha} |D^\alpha h(\xi)|^2 \rightarrow \infty$  as  $|\xi| \rightarrow \infty$ . Let  $g \in \mathcal{X}_{ac}(H)$ , the absolutely continuous subspace for  $H$ . Then there exist  $f_{\pm}$  in  $\mathcal{X}$  such that  $\|V_t g - U_t f_{\pm}\| \rightarrow 0$  as  $t \rightarrow \pm \infty$ . Consequently  $H$  on  $\mathcal{X}_{ac}(H)$  is unitarily equivalent to  $H_0$ .

*Remark 1.* If  $h(\xi) = \xi^2$  or  $h$  is elliptic or simply characteristic polynomial then the above result is known [E, Si, H2].  $h$  is said to be simply characteristic if there exists constant  $K$  such that  $\sum_{\alpha} |D^\alpha h(\xi)| \leq K[1 + |h(\xi)| + |\nabla h(\xi)|]$ .

In this article we give an outline of the proof using ideas of [Mu1, 2] of the conjecture when (i)  $n = 1$ , the function  $h$  is very general such that  $\langle Q \rangle^{-1} \langle h(P) \rangle^{-1}$  is compact and (ii)  $n \geq 2$  and  $h$  is in a large class of polynomials. For  $n = 1$ , the proof given here is extremely simpler when compared with that given in [Mu1]. Throughout this article  $K$  with or without suffix will stand for a generic constant.

**Assumption A1**  $N = \{\xi: \nabla h(\xi) = 0\}$ , the critical set, is a set of Lebesgue measure 0.

If  $h$  is a polynomial, it certainly satisfies A1.

For any real valued  $C^1$  function  $h$  on  $R^n$ , one can prove that

$$\mathcal{X}_{ac}(h(P)) = \{g \in L^2(R^n): \text{supp } \hat{g} \subseteq R^n \setminus N\}$$

where  $\hat{g}$  is the Fourier transform of  $g$ .

Thus by A1, we get  $\mathcal{X}_{ac}(H_0) = L^2(R^n)$ .

**Theorem 2.** *Let A1 hold. Then*

- (a) the wave operators  $\Omega_{\pm} = s - \lim_{t \rightarrow \pm \infty} V_t^* U_t$  exist on  $L^2(R^n)$
- (b)  $\Omega_{\pm}$  is an isometry,
- (c) (intertwining relation)  $V_t \Omega_{\pm} = \Omega_{\pm} U_t$  for all  $t$  so that  $\text{Range } \Omega_{\pm} \subseteq \mathcal{X}_{ac}(H)$  and
- (d) for  $f$  in  $\mathcal{X}_{ac}(H)$  and any real valued  $\phi$  in  $C_0^\infty(R^n \setminus N)$  we have  $s - \lim_{t \rightarrow \pm \infty} (\Omega_{\pm} - 1) \phi^2(P) V_t f = 0$ .

*Proof.* The results (a), (b), (c) are classical; we refer to, for example, [RS3]. The proof of (d) is a little bit recent [E]. When  $h(P) = P^2$ , in [E], it is proved that

$$s - \lim_{t \rightarrow \pm \infty} (\Omega_{\pm} - 1) \psi(P^2) V_t f = 0.$$

for  $f$  in  $\mathcal{X}_{ac}(H)$  and  $\psi$  in  $C_0^\infty(0, \infty)$ . The proof of (d) is similar to the proof in [E] if we replace  $\nabla P^2$  by  $\nabla h(P)$ .

To exhibit the basic physical insight in [E] for (d), we present the proof in § 2, for the sake of simplicity when  $n = 1$ . Q.E.D.

**Assumption A2.**  $C_v = h(N)$ , the set of critical values is a (countable) discrete subset of  $R$ .

When  $h$  is a polynomial, it is well known that  $C_v$  is a finite set [H2].

**Assumption A3.**  $\langle Q \rangle^{-1} \langle h(P) \rangle^{-1}$  is a compact operator.

Again, when  $h$  is a polynomial such that  $\Sigma |D^\alpha h(\xi)| \rightarrow \infty$  as  $|\xi| \rightarrow \infty$ , A3 is satisfied [Sc]. For alternate proof refer [DM].

**Lemma 3.** *Let A1, A2 and A3 hold. Let  $f \in \mathcal{X}_{ac}(H)$ . Then  $f \in \text{Range}$*

$$\Omega_{\pm} \leftrightarrow \limsup_{r \rightarrow \infty} \sup_{t \geq 0} \| \{ P^2 / (P^2 + r^2) \}^{1/2} V_t f \| = 0.$$

*Proof.*  $\Rightarrow$  (plus sign only). If  $f \in \text{Range } \Omega_+$  then there exists, by definition,  $g$  in  $\mathcal{X}$  such that  $\| V_t f - U_t g \| \rightarrow 0$  as  $t \rightarrow \infty$ . Now, the result is clear, since,  $P^2 / (P^2 + r^2)$  commutes with  $U_t$  and for any  $T$  the family  $\{ V_t f - U_t g : 0 \leq t \leq T \}$  is compact.

$\Leftarrow$  (for plus sign only). Let  $\chi$  be any "smoothened" indicator function of  $|\xi| \geq 2$ . i.e.  $\chi \in C^\infty(R^n)$ ,  $0 \leq \chi \leq 1$ ,  $\chi(\xi) = 1$  for  $|\xi| \geq 2$ , 0 for  $|\xi| \leq 1$ . Then  $\chi(\xi/r) \leq 4[\xi^2 / (\xi^2 + r^2)]^{1/2}$  gives

$$\limsup \|\chi(P/r) V_t f\| = 0. \quad (1)$$



et  $\varphi \in C_0^\infty(R \setminus C_v)$  be real valued. Then the function

$$[1 - \chi(\xi/r)]\varphi(h(\xi)) \text{ is in } C_0^\infty(R^n \setminus N).$$

by Theorem 2(d) and (1) we get

$$s - \lim_{t \rightarrow \infty} (\Omega_+ - 1)\varphi^2(H_0) V_t f = 0. \quad (2)$$

Now by A3 the operators  $(H \pm i)^{-1} - (H_0 \pm i)^{-1}$  are compact. By Stone Weierstrass theorem  $\psi(H) - \psi(H_0)$  is compact if  $\psi: R \rightarrow \mathbb{C}$  is continuous with  $\psi(\pm \infty) = 0$ . So  $H) - \varphi^2(H_0)$  is compact and so, as  $f \in \mathcal{X}_{ac}(H)$  we get

$$s - \lim_{t \rightarrow \infty} [\varphi^2(H) - \varphi^2(H_0)] V_t f = 0. \text{ By (2) we conclude}$$

$$s - \lim_{t \rightarrow \infty} V_t^* (\Omega_+ - 1) V_t \varphi^2(H) f = 0.$$

Using intertwining relation we see that  $\varphi^2(H)f \in \text{Range } \Omega_+$  for each real valued  $\varphi$  in  $(R \setminus C_v)$ . Since  $\text{Range } \Omega_+$  is closed and  $C_v$  countable we conclude that  $f \in \text{Range } \Omega_+$ .  
Q.E.D.

**Lemma 4.** (a) Let  $A$  be any Hilbert-Schmidt operator with the Hilbert-Schmidt norm  $\|A\|_2$ . Let  $H = \int \lambda dE_\lambda$  be the spectral resolution for  $H$ . Let  $f$  in  $\mathcal{X}_{ac}(H)$  be such that  $\lambda \langle E_\lambda f, f \rangle$  is bounded by  $K$  and has compact support  $S$ .  
Then

$$\int_{-\infty}^{\infty} dt \|A V_t f\|^2 \leq \|A\|_2^2 C(f)$$

where  $C(f) = K|S|^{1/2}$  depends only on  $f$ . Here  $|S|$  is the Lebesgue measure of  $S$ .  
Now let  $n = 1$ . Then

For each  $\sigma > 1/2$  and  $r \geq 1$  we have

$$\|\langle Q \rangle^{-\sigma} (P \pm ir)^{-1} \sqrt{r}\|_2^2 = K_1$$

independent of  $r$ .

Further let A3 hold. Then

$$\lim_{|k| \rightarrow \infty} \int_k^{k+1} \langle h(x) \rangle^{-2} dx = 0.$$

(improving b) For each  $\sigma > 1/2$  we have

$$\lim_{r \rightarrow \infty} \|\langle Q \rangle^{-\sigma} \sqrt{r} (P \pm ir)^{-1} (H_0 + i)^{-1}\|_2^2 = 0$$

For each  $\sigma > 1/2$  we have

*Proof.* (a) Let  $A = \sum \lambda_j \langle \cdot, \varphi_j \rangle \psi_j$  where  $\lambda_j$  is a sequence of reals decreasing to 0 with  $\|\lambda_j\|_2^2 = \sum \lambda_j^2$ ,  $\{\varphi_1, \varphi_2, \dots\}$  is ONS (orthonormal set) and  $\{\psi_1, \psi_2, \dots\}$  is also ONS. Then clearly

$$\int_{-\infty}^{\infty} dt \|AV_t f\|^2 = \sum \lambda_j^2 \int dt \left| \int d\lambda e^{-i\lambda t} \frac{d}{d\lambda} \langle E_\lambda f, \varphi_j \rangle \right|^2.$$

Now use Fourier-Plancherl Theorem to get

$$\int_{-\infty}^{\infty} dt \|AV_t f\|^2 = \sum \lambda_j^2 \int d\lambda \left| \frac{d}{d\lambda} \langle E_\lambda f, \varphi_j \rangle \right|^2. \quad (1)$$

Let  $E: \mathcal{X} \rightarrow \mathcal{X}_{ac}(H)$  be the orthogonal projection. Then for  $\varphi$  in  $\mathcal{X}$ , we easily have ([AJS], Proposition 5.18)

$$\left| \frac{d}{d\lambda} \langle E_\lambda f, \varphi \rangle \right|^2 \leq \left| \frac{d}{d\lambda} \langle E_\lambda f, f \rangle \right| \left| \frac{d}{d\lambda} \langle E_\lambda E\varphi, E\varphi \rangle \right|.$$

So by Cauchy's inequality

$$\int \left| \frac{d}{d\lambda} \langle E_\lambda f, \varphi \rangle \right|^2 d\lambda \leq K |S|^{1/2} \|E\varphi\|^2 \quad (2)$$

Using (4) in (3) we get the result:

(b) obvious.

(c) Let  $\psi$  in  $C_0^\infty(R)$  be such that  $\psi(x) = 1$  for  $|x| \leq 1$  and 0 for  $|x| \geq 2$ . Put  $\psi_k(x) = \psi(x - k)$ . Since weak limit  $(P + i)\psi_k = 0$  as  $|k| \rightarrow \infty$  and  $\langle h(Q) \rangle^{-1} (P + i)^{-1}$  is compact, the result follows.

(d) For  $r \geq 2$  we have

$$\begin{aligned} & \| \langle Q \rangle^{-\sigma} \sqrt{r} (P + ir)^{-1} (H_0 + i)^{-1} \|_2^2 \\ &= \int \langle x \rangle^{-2\sigma} dx r \int (\xi^2 + r^2)^{-1} \langle h(\xi) \rangle^{-2} d\xi \\ &\leq K_2 \sum_k r(k^2 + r^2)^{-1} \int_k^{k+1} \langle h(\xi) \rangle^{-2} d\xi. \end{aligned}$$

Now use (c) to complete the proof. (e) follows from (d) since  $(H_0 + i)(H + i)^{-1}$  is a bounded operator. Q.E.D.

*Notation* Let

$$\mathcal{L}_0^\infty(H) = \{f \in \mathcal{X}_{ac}(H) : \frac{d}{d\lambda} \langle E_\lambda f, f \rangle \text{ is bounded and has compact support}\}$$

**Theorem 5.** Let  $n = 1$  and  $A1, A2, A3$  hold.

(a) Let  $f$  be in  $\mathcal{L}_0^\infty(H)$ . Then for each  $\sigma > 1/2$ , we have

$$\lim_{r \rightarrow \infty} \int dt \| \langle Q \rangle^{-\sigma} (P \pm ir)^{-1} \sqrt{r} V_t f \|^2 = 0.$$

(b) For all  $f$  in  $\mathcal{X}_{ac}(H)$ , we have

$$\limsup_{r \rightarrow \infty} \| [P^2/(P^2 + r^2)]^{1/2} V_t f \| = 0.$$

Consequently  $\text{Range } \Omega_{\pm} = \mathcal{X}_{ac}(H)$ .

*Proof* (a). Follows from Lemma 4(a), (e) noting that  $(H + i)f$  is in  $\mathcal{L}_0^\infty(H)$ .

(b) Let  $f$  be as in (a). Write  $[A, B] = AB - BA$ . Then by the Fundamental Theorem of calculus

$$\begin{aligned} & \sup_s | \langle V_s^* P^2 (P^2 + r^2)^{-1} V_s f, f \rangle - \langle P^2 (P^2 + r^2)^{-1} f, f \rangle | \\ & \leq \int dt \left| \frac{d}{dt} \langle V_t^* P^2 (P^2 + r^2)^{-1} V_t f, f \rangle \right|. \end{aligned}$$

Now write  $P^2(P^2 + r^2)^{-1} = 1 - ir\{(P + ir)^{-1} - (P - ir)^{-1}\}$ . Note that  $[H, (P \pm ir)^{-1}] = [W(Q), (P \pm ir)^{-1}] = -(P \pm ir)^{-1} iW'(Q)(P \pm ir)^{-1}$ . Also use  $W'(Q) = \langle Q \rangle^{-\sigma} \{\langle Q \rangle^\sigma W'(Q) \langle Q \rangle^\sigma\} \langle Q \rangle^{-\sigma}$  where  $2\sigma = 1 + \delta$  and the middle term is bounded. So we get

$$\begin{aligned} & \sup_s | \langle V_s^* P^2 (P^2 + r^2)^{-1} V_s f, f \rangle - \langle P^2 (P^2 + r^2)^{-1} f, f \rangle | \\ & \leq K \int dt \| \langle Q \rangle^{-\sigma} \sqrt{r} (P \pm ir)^{-1} V_t f \|^2. \end{aligned}$$

Now the result follows for  $f$  by (a). Since  $\{f: f \text{ as in (a)}\}$  is dense in  $\mathcal{X}_{ac}(H)$  we get the result. Q.E.D.

*Remark.* A careful look at the proof of Theorem 5 shows that  $W$  need to satisfy the following weaker assumption A4 only.

*Assumption A4.*  $W$  is a real valued  $C'$  function such that for some  $\delta > 0$  we have

$$|W(x)| + |W'(x)| \leq K \langle x \rangle^{-1-\delta}.$$

For  $n \geq 2$ , a result similar to Theorem 5(a) can be got using Mourre's theory of local conjugacy [JMP, Mo].

**Theorem 6.** Let  $n \geq 2$  and the polynomial  $h$  be in any one of the following classes (a) simply characteristic polynomial, (b) parabolic polynomial i.e.  $h(\xi_1, \xi_2, \dots, \xi_n) = \xi_1 + k(\xi_2, \xi_3, \dots, \xi_n)$  (c) non-negative linear combination of non-negative monomials i.e.  $h(\xi) = \sum a_\beta \xi^\beta$  (finite sum),  $a_\beta \geq 0$ ,  $\beta = (\beta_1, \dots, \beta_n)$ ,  $\beta_j \geq 0$ ,  $\beta_j$  even for each  $j$  (d) homogeneous polynomial (e) nonisotropic (?) homogeneous polynomial i.e. there exist reals  $r_1, \dots, r_n$

and nonzero real  $m$  such that  $h(t^{r_1}\xi_1, t^{r_2}\xi_2, \dots, t^{r_n}\xi_n) = t^m h(\xi_1, \dots, \xi_n)$ . Also let  $\Sigma |D^\alpha h(\xi)| \rightarrow \infty$  as  $|\xi| \rightarrow \infty$  so that A3 holds. Then the conjecture is true for  $H = h(P) + W(Q)$ , where  $W$  is real valued, in  $C^{n+2}(R^n)$  and there exists some  $\delta > 0$  such that

$$|D^\alpha W(x)| \leq K_\alpha \langle x \rangle^{-1-\delta-|\alpha|}$$

for all  $\alpha$  with  $|\alpha| \leq n+2$

*Proof.* When  $h$  is in any of the above classes it is possible to choose

$$A = \sum Q_j \psi_j(P) + \psi_j(P) Q_j$$

where  $\psi_j$  are smooth real valued functions such that (i)  $A$  is self-adjoint, (ii) if  $H = \int \lambda dE(\lambda)$  then for each compact interval  $I \subseteq R \setminus C_v$  there exists  $\alpha > 0$  such that  $E(I) i[A, H] E(I) \geq \alpha E(I)$  (iii)  $[A, H](H+i)^{-1}$  is bounded and (iv)  $(H+i)^{-1}[A, [A, H]](H+i)^{-1}$  is bounded. Then Mourre's theory [JMP, Mo] shows that for each  $\sigma > 1/2$  and each  $f$  in  $\mathcal{H}_{ac}(H)$  with  $E(I)f = f$  we get

$$\lim_{r \rightarrow \infty} \int dt \| \langle Q \rangle^{-\sigma} \langle P \rangle^{1/2} r(P^2 + r^2)^{-1} V_t f \|^2 = 0. \quad (5)$$

Compare this result with Theorem 5(a). Now from (5) we easily get

$$\limsup_{r \rightarrow \infty} \int_t \| \{ \langle P \rangle^2 / (P^2 + r^2) \}^{1/2} V_t f \| = 0$$

for each  $f$  in  $\mathcal{H}_{ac}(H)$ . Now the result follows from Lemma 3.

Q.E.D.

## 2. Proof of Theorem 2

Theorem 2(a) is a consequence of "Propagation estimate-1 (PE1)" of Lemma 8 while Theorem 2(d) requires PE2 from Lemma 10. Let  $n = 1$  and put

$$G = \{ \xi : h'(\xi) \neq 0 \} = R \setminus N$$

$\chi(A)$  will stand for the indicator function of the set  $A$ .

*Lemma 7.* Let  $\varphi \in C_0^\infty(G)$  and  $3a = \inf \{ |h'(\xi)| : \xi \in \text{supp } \varphi \}$ , then  $\| \chi(|Q| \leq a|t|) U_t \varphi(P) \chi(|Q| \leq a|t|) \|_2 \leq K_2(\varphi) \langle t \rangle^{-2}$  for all real  $t$ . Here  $K_2(\varphi)$  depends on finitely many derivatives of  $\varphi$  and  $h$ .

*Proof.* We prove the result for  $t \geq 0$ . We need to prove it only for  $t \geq 1$ , since for  $t \leq 1$  the result is clear. Put

$$I(t) = \chi(|Q| \leq at) U_t \varphi(P) \chi(|Q| \leq at).$$

Then  $I(t)$  is an integral operator with integral kernel  $I(t, q, x)$  given by

$$I(t, q, x) = \chi(|q| \leq at) \chi(|x| \leq at) \int d\xi \varphi(\xi) \exp(i[q\xi - x\xi - th(\xi)]) \quad (6)$$

where, of course,  $[I(t)f](q) = \int I(t, q, x)f(x)dx$ . We apply the method of (non) stationary phase. When  $|q| \leq at$ ,  $|x| \leq at$  and  $\xi$  is in  $\text{supp } \varphi$  one easily sees that

$$|q - x - th'(\xi)| \geq t|h'(\xi)| - |q| - |x| \geq at. \quad (7)$$

Also we have

$$\begin{aligned} \exp(i[q\xi - x\xi - th(\xi)]) &= -i\{q - x - th'(\xi)\}^{-1} \frac{d}{d\xi} \\ &\quad \exp(i[q\xi - x\xi - th(\xi)]). \end{aligned} \quad (8)$$

Using (8) in (6) and integrating by parts once we get

$$\begin{aligned} I(t, q, x) &= i\chi(|q| \leq at)\chi(|x| \leq at) \int d\xi \exp(i[q\xi - x\xi - th(\xi)]). \\ &\quad \varphi'(\xi)[q - x - th'(\xi)]^{-1} + \varphi(\xi)[q - x - th'(\xi)]^{-2} th''(\xi). \end{aligned} \quad (9)$$

So by (7) we get

$$\begin{aligned} |I(t, q, x)| &\leq \chi(|q| \leq at)\chi(|x| \leq at) \sup_{\xi \in \text{supp } \varphi} \\ &\quad \left\{ |q - x - th'(\xi)|^{-1} \int |\varphi'(u)| du + |q - x - th'(\xi)|^{-2} t \int |h''(u)\varphi(u)| du \right\} \\ &\leq \chi(|q| \leq at)\chi(|x| \leq at)(at)^{-1} K_1(\varphi) \end{aligned} \quad (10)$$

Here  $K_1(\varphi)$  depends on  $\varphi, \varphi', h'$  and  $h''$ .

Now use (8) in (9) integrate by parts, get an identity. Again use (8) in (9), integrate by parts, and continue the process. Just as we got the inequality (10) from the identity (9), we can get for each  $N$

$$|I(t, q, x)| \leq \chi(|q| \leq at)\chi(|x| \leq at)(at)^{-N} K_N(\varphi)$$

where  $K_N(\varphi)$  depends only on finitely many derivatives of  $\varphi$  and  $h$ . Now it is clear that  $\|I(t)\|_2 \leq K_N(\varphi)t^{-N+1}$ . Take  $N=3$  to get the result. Q.E.D.

**Lemma 8.** (Propagation estimate 1), Let  $\varphi \in C_0^\infty(G)$ . Then

$$\|\langle Q \rangle^{-1-\delta} U_t \varphi(P) \langle Q \rangle^{-1-\delta}\| \leq K(\varphi) \langle t \rangle^{-1-\delta} \text{ for } 0 \leq \delta \leq 1.$$

*Proof.* Let  $\chi = \chi(|Q| \leq a|t|)$  and put  $\sigma = 1 + \delta$ . The result follows from Lemma 7 by using the identity

$$\begin{aligned} \langle Q \rangle^{-\sigma} U_t \varphi(P) \langle Q \rangle^{-\sigma} &= \langle Q \rangle^{-\sigma} \chi U_t \varphi(P) \chi \langle Q \rangle^{-\sigma} + \langle Q \rangle^{-\sigma} (1 - \chi) \\ &\quad U_t \varphi(P) \chi \langle Q \rangle^{-\sigma} + \langle Q \rangle^{-\sigma} U_t \varphi(P) (1 - \chi) \langle Q \rangle^{-\sigma} \end{aligned}$$

and observing that

$$\|\langle Q \rangle^{-\sigma} (1 - \chi)\| = \langle at \rangle^{-\sigma} = \|(1 - \chi) \langle Q \rangle^{-\sigma}\| \quad \text{Q.E.D.}$$

*Proof of Theorem 2a* (for the plus sign only). Since  $R - G = N$  has zero Lebesgue

is dense in  $L^2(R)$ . Here  $\mathcal{S}(R)$  stands for the Schwartz space of smooth rapidly decreasing functions on  $R$ . For any  $f$  in  $\mathcal{D}$ , choose  $\varphi$  in  $C_0^\infty(G)$  so that  $\varphi = 1$  on  $\text{supp } \hat{f}$ . This gives  $\varphi(P)f = f$ . Now

$$\begin{aligned} \|V_t^* U_t f - V_s^* U_s f\| &= \left\| \int_s^t du \frac{d}{du} V_u^* U_u \varphi(P) f \right\| \\ &= \left\| \int_s^t du V_u^* W(Q) \varphi(P) U_u \langle Q \rangle^{-1-\delta} \langle Q \rangle^{1+\delta} f \right\| \\ &\leq \int_s^t du \| \langle Q \rangle^{-1-\delta} U_u \varphi(P) \langle Q \rangle^{-1-\delta} \| \| \langle Q \rangle^{1+\delta} f \|. \end{aligned}$$

By Lemma 8 we see that  $\text{RHS} \rightarrow 0$  as  $s, t \rightarrow \infty$ . So  $V_t^* U_t f$  has limit as  $t \rightarrow \infty$ . The result follows since  $\mathcal{D}$  is dense and  $\{V_t^* U_t; t \geq 0\}$  is norm bounded. Q.E.D.

*Proof of Theorem 2b* is obvious since  $V_t^* U_t$  is an isometry for each  $t$ . Q.E.D.

*Proof of Theorem 2c* (for the plus sign only). For  $x$  in  $\mathcal{X}$ ,

$$\begin{aligned} V_s \Omega_+ x &= V_s \lim_{t \rightarrow \infty} V_t^* U_t x = \lim_{t-s \rightarrow \infty} V_{t-s}^* U_{t-s} U_s x \\ &= \Omega_+ U_s x \end{aligned}$$

proving the intertwining relation.

For  $\varphi$  in  $\mathcal{S}(R)$  we have  $\varphi(H)\Omega_+ x = \int ds \hat{\varphi}(s) e^{isH} \Omega_+ x$  and

$$\Omega_+ \varphi(H_0) x = \int ds \hat{\varphi}(s) \Omega_+ e^{isH_0} x.$$

Using the intertwining relation we get  $\varphi(H)\Omega_+ = \Omega_+ \varphi(H_0)$ . If  $H = \int \lambda dE(\lambda)$  and  $H_0 = \int \lambda dE_0(\lambda)$  are the spectral representations for  $H$  and  $H_0$  then one easily sees that for each Borel subset  $A$  of  $R$

$$E(A)\Omega_+ = \Omega_+ E_0(A). \quad (11)$$

For any  $x$  in  $\mathcal{X}$ , using (11) and Theorem 2(b) we have

$$\langle E(A)\Omega_+ x, \Omega_+ x \rangle = \|E(A)\Omega_+ x\|^2 = \langle E_0(A)x, x \rangle$$

Now the result is clear since  $\mathcal{X}_{ac}(H_0) = L^2(R)$ . Q.E.D.

*Lemma 9.* Let  $\varphi$  and  $a$  be as in Lemma 7. Then (a) and (b) hold for  $t \geq 0$ , (c) and (d) hold for  $t \leq 0$ .

$$(a) \|\chi(|Q| \leq a|t|) U_t \varphi(P) \chi(h'(P) \geq 0) \chi(Q \geq 0)\|_2 \leq K(\varphi) \langle t \rangle^{-2}$$

$$(b) \|\chi(|Q| \leq a|t|) U_t \varphi(P) \chi(h'(P) \leq 0) \chi(Q \leq 0)\|_2 \leq K(\varphi) \langle t \rangle^{-2}$$

$$(c) \|\chi(|Q| \leq a|t|) U_t \varphi(P) \chi(h'(P) \geq 0) \chi(Q \leq 0)\|_2 \leq K(\varphi) \langle t \rangle^{-2}$$

$$(d) \|\chi(|Q| \leq a|t|) U_t \varphi(P) \chi(h'(P) \leq 0) \chi(Q \geq 0)\|_2 \leq K(\varphi) \langle t \rangle^{-2}$$

*Proof* (a) Similar to the proof of Lemma 7. One notes that if  $h'(\xi) \geq 0$ ,  $\xi$  is in  $\text{supp } \varphi$  and  $x \geq 0$  then  $|x + th'(\xi)| \geq 3a|t| + |x|$ . If in addition  $|q| \leq at$ , then  $|q - x - th'(\xi)| \geq |x| + 2a|t|$ .

The proofs of (b), (c), (d) are similar to that of (a)

Q.E.D.

**Lemma 10.** (Propagation estimate-2). Let  $F_+$  and  $F_-$  be the non-negative self-adjoint operators given by

$$F_+ = \chi(h'(P) \geq 0) \chi(Q \geq 0) \chi(h'(P) \geq 0) + \chi(h'(P) \leq 0)$$

$$\chi(Q \leq 0) \chi(h'(P) \leq 0)$$

$$F_- = \chi(h'(P) \geq 0) \chi(Q \leq 0) \chi(h'(P) \geq 0) + \chi(h'(P) \leq 0)$$

$$\chi(Q \geq 0) \chi(h'(P) \leq 0)$$

so that  $F_+ + F_- = 1$ . Then for  $\varphi$  in  $C_0^\infty(G)$  we have

$$\|\langle Q \rangle^{-1-\delta} U_t \varphi(P) F_\pm\| \leq K(\varphi) \langle t \rangle^{-1-\delta} \text{ for } t \geq 0 \text{ and } 0 \leq \delta \leq 1.$$

*Proof.* Similar to the proof of Lemma 8 using Lemma 9.

Q.E.D.

**Lemma 11.** Let  $\varphi$  in  $C_0^\infty(G)$  be real valued. Then

(a)  $(\Omega_\pm - 1) \varphi(P) F_\pm$  is compact.

(b)  $\lim_{t \rightarrow \pm \infty} F_\pm \varphi(P) U_{-t} g = 0$  for all  $g$  in  $L^2(R)$

*Proof.* (For plus sign only) (a) clearly

$$(\Omega_+ - 1) \varphi(P) F_+ = i \int_0^\infty dt V_t^* W(Q) \varphi(P) U_t F_+$$

Now  $W(Q) \varphi(P)$  is compact and by Lemma 10 we get

$$\int_0^\infty dt \|W(Q) \varphi(P) U_t F_+\| < \infty.$$

The result is clear.

(b) By Lemma 10,

$$\lim_{t \rightarrow \infty} \|F_+ \varphi(P) U_{-t} \langle Q \rangle^{-1-\delta}\| = 0$$

so that for  $g$  in  $\mathcal{S}(R)$  we have

$$\lim_{t \rightarrow \infty} F_+ \varphi(P) U_{-t} g = 0.$$

The result follows since  $\mathcal{S}(R)$  is dense in  $L^2(R)$ .

Q.E.D.

*Proof of Theorem 2d.* Let  $\varphi$  in  $C_0^\infty(G)$  be real valued. By Lemma 11 (a),  $(\Omega_+ - 1)\varphi(P)F_+ \varphi(P)$  is compact. Since  $\lim_{t \rightarrow \infty} V_t f = 0$  as  $t \rightarrow \infty$  we have

$$\lim_{t \rightarrow \infty} (\Omega_+ - 1)\varphi(P)F_+ \varphi(P) V_t f = 0. \quad (12)$$

Similarly we have  $\lim_{t \rightarrow \infty} [(\Omega_- - 1)\varphi(P)F_- \varphi(P)]^* V_t f = 0$  proving

$$\lim_{t \rightarrow \infty} \varphi(P)F_- \varphi(P) V_t f - \varphi(P)F_- \varphi(P)\Omega_-^* V_t f = 0. \quad (13)$$

Taking adjoint for the intertwining relation we get  $\Omega_-^* V_t = U_t \Omega_-^*$ . So by (13) and Lemma 11(b) we conclude  $\lim_{t \rightarrow \infty} \varphi(P)F_- \varphi(P) V_t f = 0$ . So, easily

$$\lim_{t \rightarrow \infty} (\Omega_+ - 1)\varphi(P)F_- \varphi(P) V_t f = 0. \quad (14)$$

We get the result by adding (12) with (14) and noting  $F_+ + F_- = 1$ .

Q.E.D.

## References

- [AJS] Amrein W O, Jauch J M and Sinha K B, *Scattering Theory in Quantum Mechanics* (Lecture notes and Suppl in Physics 16) (Reading, Benjamin) (1977)
- [DM] Davies E B and Muthuramalingam P L, Trace properties of some highly anisotropic operators, *J. London Math. Soc.* **31** (1985) 137–149
- [E] Enss V, Asymptotic completeness of quantum mechanical potential scattering I. Short range potentials, *Commun. Math. Phys.* **61** (1978) 285–291
- [H2] Hormander L, *The analysis of linear partial differential operators; Vol 2. Differential operators with constant coefficients* (Berlin, Springer Verlag) (1983)
- [JMP] Jensen A, Mourre E and Perry P, Multiple commutator estimates and resolvent smoothness in scattering theory, *Ann. Inst. H. Poincaré* **41** (1984) 207–225
- [Mo] Mourre E, Absence of singular continuous spectrum for certain self-adjoint operators *Comm. Math. Phys.* **78** (1981) 391–408
- [Mu1] Muthuramalingam P L, Bound states for momentum and asymptotic completeness in  $L^2(R^n)$ : Trace class commutators for  $n = 1$ , *Rev. Roum. Math. Pures. Appl.* **37** (1992) 747–761
- [Mu2] Muthuramalingam P L, Bound states for momentum and asymptotic completeness in  $L^2(R^n)$ : II. Mourre's theory of local conjugacy for  $n \geq 2$ , *J. Fac. Sci. Univ. Tokyo. IA Math.* **39** (1992) 185–205
- [RS3] Reed M and Simon B, *Methods of Modern Mathematical Physics Vol 3: Scattering theory* (New York, Academic Press) (1979)
- [Sc] Schechter M, *Spectra of partial differential operators* (Amsterdam: NH Publishing Co.) (1971)
- [Si] Simon B, Phase space analysis of simple scattering systems: extensions of some work of Enss, *Duke. Math. J.* **46** (1979) 119–168



## The geometry and spectra of hyperbolic manifolds\*

PETER D HISLOP

Mathematics Department, University of Kentucky, Lexington, KY 40506-0027, USA

**Abstract.** This paper is a self-contained discussion of the relationship between spectral and geometric properties of a class of hyperbolic manifolds. After a review of the fundamentals of hyperbolic manifolds, aspects of the theory for the compact case and the finite-volume case are discussed. The main emphasis of this work is on a class of infinite-volume hyperbolic manifolds  $\mathcal{M}$  which arise as quotients of hyperbolic space  $H^n$  by discrete subgroups  $\Gamma$ , i.e.  $\mathcal{M} \cong H^n/\Gamma$ . This paper describes joint work with R G Froese and P A Perry. For these infinite-volume hyperbolic manifolds, there are very few eigenvalues, so most of the spectral information is carried by the generalized eigenfunctions of the Laplacian. These eigenfunctions can be constructed from the asymptotics of the Green's function. It is shown how the asymptotic geometry of the manifold determines the asymptotic behavior of the Green's function, and hence the eigenfunctions, near infinity. This information is used to construct an  $S$ -matrix for the manifold which is a pseudo-differential operator acting on sections of a fibre bundle over the boundary of the manifold at infinity. The meromorphic properties of this operator and its inverse, as a function of the spectral parameter, are described. A functional relation between the  $S$ -matrix and the generalized eigenfunctions is derived. An important consequence of this relation and the meromorphicity of the  $S$ -matrix and its inverse is the existence of the meromorphic continuation of the Eisenstein series associated with the discrete group  $\Gamma$ . Finally, an overview of recent progress and some open problems are presented, including a discussion of the asymptotic behavior of the counting function for the scattering poles.

**Keywords.** Hyperbolic manifolds; Laplacians; Eisenstein series.

### Introduction

This paper is an expanded version of the lectures which I gave at the Summer Workshop on Spectral and Inverse Spectral Theories at Kodaikanal, India during August 24–30, 1993. The results in chapters 2–4 is joint work with R G Froese of the University of British Columbia; Vancouver, Canada and P A Perry of the University of Kentucky, Lexington, KY. I would like to thank them for such a fruitful and enjoyable collaboration. I would also like to thank K Pinney Mortensen and R Molzon for some discussions on differential geometry.

The goal of these lectures is to provide a rather complete exposition of the spectral theory of a class of infinite volume hyperbolic manifolds called geometrically finite. The emphasis is on the relationship between the analysis and the geometry of these manifolds. The first chapter is a summary of the parts of the theory of hyperbolic

---

\* Research supported in part by NSF grant DMS93–07438

compact hyperbolic manifolds. There are many very good papers on this case. The technique of local positive commutators is given: it seems to be rather economical and complete. Chapters 3 and 4 are the core of the lectures. The Green's function, Eisenstein series and  $S$ -matrix are discussed in some detail. Finally, Chapter 5 presents an overview of many areas of current research.

## 1. Hyperbolic manifolds

The aim of this chapter is to define a class of hyperbolic manifolds particularly amenable to spectral analysis. To achieve this, we will discuss the basic geometric and group theoretic properties of these manifolds. We begin with the unique simply connected  $n$ -dimensional hyperbolic manifold  $\mathbb{H}^n$  and its models. We then discuss the isometry group of  $\mathbb{H}^n$  and discrete subgroups. Finally, we present a uniformization theorem which describes arbitrary hyperbolic manifolds. The main results in this chapter are standard. We refer to the books [A1], [B], [BeP], [Ca], [T], and [Th] for further details concerning differential geometry and the geometry of hyperbolic manifolds.

### 1.1 Models for hyperbolic space

#### DEFINITION 1.1

The  $n$ -dimensional hyperbolic space  $\mathbb{H}^n$  is a complete, simply connected Riemannian manifold with all sectional curvatures equal to  $-1$ .

Hence, hyperbolic space  $\mathbb{H}^n$  has constant negative scalar curvature equal to  $-n < 0$ . There are three standard models for  $\mathbb{H}^n$  which are, of course, all isometric. These models are useful for computations and demonstrate that the definition above is not an empty one.

(1) *Upper half space.*  $\mathbb{H}^n$  is realized as a set

$$U_n \equiv \mathbb{R}^{n-1} \times \mathbb{R}_+ = \{x \in \mathbb{R}^n \mid x_n > 0, \quad x' \equiv \{x_1, \dots, x_{n-1}\} \in \mathbb{R}^{n-1}\}.$$

We take  $x = (x_1, \dots, x_n)$  as global coordinates and write  $x = (x', x_n) \in \mathbb{R}^{n-1} \times \mathbb{R}_+$ . The metric in these coordinates is

$$ds^2 = x_n^{-2} \left( \sum_{i=1}^n dx_i^2 \right) = \sum_{i,j=1}^n g_{ij}(x) dx_i dx_j, \quad (1.1)$$

and

$$g_{U_n}(x) = x_n^{-2} \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \quad (1.2)$$

Note that the metric degenerates on  $\mathbb{R}^{n-1} \cup \{\infty\}$ , where  $\mathbb{R}^{n-1}$  is interpreted as the hyperplane  $x_n = 0$ . The hyperbolic distance  $d_h$  between  $z, w \in \mathbb{H}^n$  is defined through

$$\sigma(z, w) = (1/2)(1 + \cosh d_h(z, w)) = (\|z - w\|_E + (z_n + w_n)^-)(4z_n w_n)^{-1/2}, \quad (1.3)$$

where  $\|x\|_E^2 = \sum_{i=1}^n |x_i|^2$  denotes the Euclidean norm of  $x$ .

(2) *Ball model.*  $\mathbb{H}^n$  is realized as the set

$$B_n \equiv \{x \in \mathbb{R}^n \mid \|x\|_E < 1\}.$$

We take  $x$  as a global coordinate and define a metric

$$ds^2 = 4(1 - \|x\|_E^2)^{-2} \left( \sum_{i=1}^n dx_i^2 \right). \quad (1.4)$$

We call  $\{x \mid \|x\|_E = 1\} \equiv S^{n-1}$ , the boundary at infinity  $\partial_\infty B_n$ . Note that the metric  $g_{B_n}$  degenerates on  $S^{n-1}$ . Comparing with  $U_n$ , we see that whereas  $\partial_\infty U_n = \mathbb{R}^{n-1} \cup \{\infty\}$  has an apparent "distinguished" point at infinity, this is not really the case as the ball model shows. Hence, the boundary at infinity of  $\mathbb{H}^n$ ,  $\partial_\infty \mathbb{H}^n$ , is the one point compactification of  $\mathbb{R}^{n-1}$ , i.e.  $S^{n-1}$ . In the model  $B_n$ , the distance function is

$$\cosh(d_h(x, y)) = 1 + \frac{2\|x - y\|_E^2}{(1 - \|x\|_E^2)(1 - \|y\|_E^2)}. \quad (1.5)$$

The relation between  $U_n$  and  $B_n$  is given by fractional linear transformations:  $G: B_n \rightarrow U_n$  is defined by

$$G(x) = (1 + \|x\|_E^2 - 2x_1)^{-1} (x_1, x_2, \dots, \frac{1}{2}(1 - \|x\|_E^2)). \quad (1.6)$$

The inverse map is given by  $G^{-1}: U_n \rightarrow B_n$ ,

$$G^{-1}(z) = ((z_n + \frac{1}{2})^2 + \|z'\|_E^2)^{-1} (z_1, \dots, z_{n-1}, z_n^2 + \|z'\|_E^2 - \frac{1}{4}). \quad (1.7)$$

When  $n=2$ , these reduce to the usual conformal transformations. The map  $G$  is a diffeomorphism and an isometry between manifolds, i.e. if  $x \in B_n$  and  $v, w \in T_x B_n$ , and  $G_*^{-1}: TB_n \rightarrow TU_n$ , then

$$g_{B_n}(x)(v, w) = g_{U_n}(G(x))(G_*^{-1}v, G_*^{-1}w).$$

(3) *Hyperboloid model.*  $\mathbb{H}^n$  is realized as a subset of  $\mathbb{R}^{n+1}$  determined by the indefinite quadratic form  $q: \mathbb{R}^{n+1} \times \mathbb{R}^{n+1} \rightarrow \mathbb{R}^+ \cup \{0\}$  given by

$$q(x, y) \equiv x_{n+1}y_{n+1} - \sum_{i=1}^n x_i y_i. \quad (1.8)$$

The group of linear transformations preserving  $q$  is denoted by  $O(n, 1)$ . We consider the surface

$$Q_n^+ \equiv \{x \in \mathbb{R}^{n+1} \mid q(x, x) = 1 \text{ and } x_{n+1} > 0\}. \quad (1.9)$$

Note that  $Q_n^+$  is the positive sheet of the two-sheeted hyperboloid in  $\mathbb{R}^{n+1}$  determined

by  $q(x, x) = 1$  and has co-dimension 1. It is easy to check that  $K \in O(n, 1)$  preserves  $Q_n^+$  if and only if  $K_{n+1, n+1} > 0$ . We introduce a metric on  $Q_n^+$  by setting

$$ds^2 = \sum_{i=1}^n dx_i^2 - dx_{n+1}^2. \quad (1.10)$$

To verify that this is a metric, we must prove that it defines a positive definite inner product on each  $T_x Q_n^+$ . Let  $\gamma: [0, 1] \rightarrow Q_n^+$  be a  $C^1$ -curve;  $\dot{\gamma}(t) \in T_{\gamma(t)} Q_n^+$ . We compute the form  $q(\dot{\gamma}, \dot{\gamma})$ , identifying  $T_{\gamma(t)} Q_n^+$  as a subset of  $\mathbb{R}^{n+1}$ . Since  $q(\gamma, \gamma) = 1$ , we have  $q(\dot{\gamma}, \gamma) = 0$  so

$$\begin{aligned} q(\dot{\gamma}, \dot{\gamma}) &= \left( \frac{\dot{\gamma}_{n+1} \gamma_{n+1}}{\gamma_{n+1}} \right)^2 - \sum_{i=1}^n \dot{\gamma}_i^2 \\ &= \frac{\left( \sum_{i=1}^n \gamma_i \dot{\gamma}_i \right)^2}{\gamma_{n+1}^2} - \sum_{i=1}^n \dot{\gamma}_i^2 \\ &\leq \left( \sum_{i=1}^n \dot{\gamma}_i^2 \right) \left( \frac{\sum_{i=1}^n \gamma_i^2}{\gamma_{n+1}^2} - 1 \right) \\ &= - \frac{\sum_{i=1}^n \dot{\gamma}_i^2}{\gamma_{n+1}^2} \leq 0. \end{aligned}$$

If  $q(\dot{\gamma}, \dot{\gamma}) = 0$  then  $\dot{\gamma}_i = 0$ ,  $i = 1, \dots, n$  so  $\dot{\gamma}_{n+1} = 0$  i.e.  $-q$  is positive definite on  $Q_n^+$ . We can define a distance function for  $x, y \in Q_n^+$ , by

$$d_h(x, y) \equiv \inf_{\substack{\gamma \in Q_n^+ \\ \gamma(0)=x, \gamma(1)=y}} \int_0^1 [-q(\dot{\gamma}, \dot{\gamma})]^{1/2} dt, \quad (1.11)$$

for which the corresponding metric is given above. Note that  $(x_1, \dots, x_{n+1})$  provides a global parametrization for  $Q_n^+$ . It is not hard to compute the curvatures of  $Q_n^+$  with this metric and show that they are  $-1$ . Consequently,  $Q_n^+$  is equivalent (isometric) to  $U_n$  and  $B_n$ . For example,  $F: Q_n^+ \rightarrow B_n$ , is given by the standard projective map

$$F(x_1, \dots, x_{n+1}) = \left( \frac{x_1}{1 + x_{n+1}}, \dots, \frac{x_n}{1 + x_{n+1}} \right). \quad (1.12)$$

The inverse map  $F^{-1}: B_n \rightarrow Q_n^+$  is given by

$$F^{-1}(y) = \left( \frac{2y_1}{1 - \|y\|_E}, \frac{2y_2}{1 - \|y\|_E}, \dots, \frac{1 + \|y\|_E}{1 - \|y\|_E} \right). \quad (1.13)$$

The boundary at infinity is the "circle" obtained from the  $F^{-1}$  map as  $\|y\| \rightarrow 1$ .

In these lectures, we will always work with the upper half space model  $U_n$ . Henceforth we will write  $\mathbb{H}^n$  for this model. However, all of the calculations can be done in any model. We will make use of  $Q_n^+$  when we discuss hyperbolic isometries below. It is sometimes easier to visualize  $\partial_\infty \mathbb{H}^n$  by looking at the ball model  $B_n$ .

We need one other fact about the Riemannian geometry of  $\mathbb{H}^n$ . Suppose  $\gamma: [0, 1] \rightarrow \mathcal{M}$  is a  $C^1$ -curve on a Riemannian manifold  $\mathcal{M}$ . The tangent vector  $\dot{\gamma}(t) \in T_{\gamma(t)}\mathcal{M}$ . We define  $\|\dot{\gamma}(t)\|_g^2 = \langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)}$ , where  $\langle \cdot, \cdot \rangle$  is the inner product on  $T_{\gamma(t)}\mathcal{M}$  associated with the metric  $g$ . We can define the length of  $\gamma$ ,  $L(\gamma)$ , by

$$L(\gamma) = \int_0^1 \|\dot{\gamma}(t)\|_g dt = \int_0^1 \left[ \sum_{i,j} g_{ij}(\gamma(t)) \dot{\gamma}_i(t) \dot{\gamma}_j(t) \right]^{1/2} dt. \quad (1.14)$$

Specializing to  $\mathbb{H}^n$ , we obtain

$$L(\gamma) = \int_0^1 \gamma_n(t)^{-1} \|\dot{\gamma}(t)\|_E dt. \quad (1.15)$$

We can extend the functional  $L$  to piecewise continuous curves,  $\gamma \in PC[0, 1]$ , in an obvious manner. For a given length function  $L$ , we define the distance between two points  $p, q \in \mathcal{M}$  by

$$d(p, q) \equiv \inf \{ L(\gamma) \mid \gamma \in PC[0, 1], \gamma(0) = p \text{ and } \gamma(1) = q \}.$$

## DEFINITION 1.2

A geodesic on a Riemannian manifold  $(\mathcal{M}, g)$  is a curve  $\gamma \in PC[0, 1]$  which is an extremal of the length function  $L$  (1.14).

Recall that an extremal of  $L$  is a curve  $\gamma$  such that the first variation  $\delta L / \delta \gamma = 0$ . For  $\gamma$  to be a geodesic it must satisfy, in each local coordinate chart, the differential equation

$$\frac{d^2 \gamma^i(s)}{ds^2} + \sum_{j,k=1}^n \Gamma_{jk}^i(\gamma) \frac{d\gamma^j}{ds} \frac{d\gamma^k}{ds} = 0, \quad (1.16)$$

where

$$\Gamma_{ij}^k = \frac{1}{2} \sum_{l=1}^n g^{kl} \left( \frac{\partial g_{il}}{\partial x_j} + \frac{\partial g_{jl}}{\partial x_i} - \frac{\partial g_{ij}}{\partial x_l} \right),$$

and  $g^{kl}$  is  $[g^{-1}]_{kl}$ . For the hyperbolic metric  $g$  in (1.2), we easily find

$$\Gamma_{ij}^k = -x_n^{-1} [\delta_{ik} \delta_{jn} + \delta_{jk} \delta_{in} - \delta_{kn} \delta_{ij}].$$

To solve the equations (1.16) on  $\mathbb{H}^n$ , we recall that the solutions are parametrized by arc length. This is the parametrization for which

$$s(t) = \int_0^t \|\dot{\gamma}(u)\|_g du = t. \quad (1.17)$$

Consider the  $C^\infty$ -curve in  $\mathbb{H}^n$  given by

$$\gamma(s) = \begin{cases} \gamma^n(s) = s \\ \gamma^i(s) = 0, i = 1, \dots, n-1, \end{cases}$$

for  $s > 0$ . This curve satisfies (1.15)–(1.17) and hence is a geodesic. If we fix  $e_n \equiv (0, \dots,$

$0, 1) \in \mathbb{H}^n$ , then the hyperbolic distance from  $e_n$  to  $te_n$ ,  $t > 0$ , along  $\gamma$  is

$$d_h(e_n, te_n) = |\log t|.$$

This shows that the Euclidean length of the geodesic  $\gamma$  grows exponentially in the hyperbolic distance from  $e_n$ . It is easy to see that any vertical line

$$\gamma(t) = te_n + (a_1, \dots, a_{n-1}, 0), t > 0, \quad (1.18)$$

is geodesic. If  $\sigma: \mathbb{H}^n \rightarrow \mathbb{H}^n$  is a diffeomorphism preserving the metric and  $\gamma(t)$  is a geodesic, then  $\sigma \circ \gamma(t)$  is also a geodesic. It follows from Corollary 1.10 of the next section that all such isometries of  $\mathbb{H}^n$  are conformal maps. Consequently, we obtain

### PROPOSITION 1.3

*The geodesics of  $\mathbb{H}^n$  are Euclidean semi-circles with centers on  $\partial_\infty \mathbb{H}^n$ . If the center is at  $\infty$ , then the geodesics are the straight lines (1.18).*

Note that this proposition implies that  $\mathbb{H}^n$  is geodesically complete, i.e. that any geodesic can be infinitely extended. This follows from the fact that any geodesic of the form (1.18) can be infinitely extended and any geodesic is of the form  $\sigma \circ \gamma$ ,  $\sigma$  an isometry of  $\mathbb{H}^n$  and  $\gamma$  as in (1.18). The Rinow–Hopf Theorem (cf. [Ca]) states that a Riemannian manifold is complete if and only if it is geodesically complete. Hence,  $\mathbb{H}^n$  is complete.

Geodesics have been defined as extremals of the length functional, but they might not be minimizers. However, for the manifolds of interest to us, this is the case.

### PROPOSITION 1.4

*If  $\mathcal{M}$  is a complete Riemannian manifold then any two points can be joined by a geodesic which minimizes the length function i.e. for  $x, y \in \mathcal{M} \exists \gamma: [0, 1] \rightarrow \mathcal{M}$  s.t.  $\gamma(0) = x, \gamma(1) = y$ ,  $\gamma$  is geodesic and  $L(\gamma) = d_{\mathcal{M}}(x, y)$ . If in addition, the exponential map  $\exp_x: T_x \mathcal{M} \rightarrow \mathcal{M}$  is 1:1, such a minimizing geodesic is unique.*

### COROLLARY 1.5

*Any two points  $x, y \in \mathbb{H}^n$  are connected by a unique geodesic such that  $L(\gamma) = d_h(x, y)$ .*

The proof of the first result follows from the Rinow–Hopf Theorem. The uniqueness of the minimizer is clear if the exponential map is 1:1. The injectivity of the exponential map for  $\mathbb{H}^n$  (at any point) is a consequence of the Cartan–Hadamard Theorem. This states that a complete, simply-connected,  $n$ -dimensional Riemannian manifold of non-positive sectional curvature is diffeomorphic to  $\mathbb{R}^n$ , i.e. the exponential map is 1:1 (see [Ca] for proofs of these facts).

## 1.3 Isometry Group of $\mathbb{H}^n$

Let  $(\mathcal{M}, g)$  be a Riemannian manifold. We are interested in smooth maps  $\phi: \mathcal{M} \rightarrow \mathcal{M}$  which preserve the metric.

### DEFINITION 1.6

The isometry group of  $(\mathcal{M}, g)$ ,  $\text{Isom } \mathcal{M}$ , is the group of orientation preserving diffeomorphisms  $\phi: \mathcal{M} \rightarrow \mathcal{M}$  such that if  $\phi_*: TM \rightarrow TM$  is the induced map, then for

$$x \in \mathcal{M}, v, w \in T_x \mathcal{M},$$

$$\langle \phi_* v, \phi_* w \rangle_{g(\phi(x))} = \langle v, w \rangle_{g(x)}. \quad (1.19)$$

An immediate consequence of this definition is that for all  $x, y \in \mathcal{M}$ ,

$$d_g(\phi(x), \phi(y)) = d_g(x, y). \quad (1.20)$$

Infinitesimally, condition (1.19) means that

$$ds'^2 = \sum_i g_{ij}(x') dx'_i dx'_j = \sum_{ij} g_{ij}(x) dx_i dx_j = ds^2, \quad (1.21)$$

where  $\phi(x) \equiv x'$ . We wish to identify  $\text{Isom } \mathbb{H}^n$  as a classical Lie group. It is known that for a general Riemannian manifold  $\mathcal{M}$ ,  $\text{Isom } \mathcal{M}$  is a real Lie group.

$\mathbb{H}^2$ . We can write the metric as  $ds^2 = x_2^{-2}(dx_1^2 + dx_2^2)$ . Clearly any Euclidean transformation in the hyperplane (for  $n=2$ , a line)  $x_2 = c$ ,  $c > 0$  constant, is an isometry. These are translations in  $x_1$ :

$$\phi_T(x_1, x_2) = (x_1 + b, x_2), \text{ any } b \in \mathbb{R}. \quad (1.22)$$

We also note that the metric is homogeneous of order zero and hence invariant under scale transformations or dilations,

$$\phi_D(x_1, x_2) = (\lambda x_1, \lambda x_2), \quad \lambda > 0. \quad (1.23)$$

But there are other isometries, for example, the inversion in the unit semi-circle

$$\phi_I(x_1, x_2) = \left( -\frac{x_1}{\|x\|_E^2}, \frac{x_2}{\|x\|_E^2} \right). \quad (1.24)$$

To verify this, we can compute the differentials. However, an easier way is to identify  $\mathbb{H}^2$  with  $\mathbb{C}^+ = \{z = x_1 + ix_2 | x_1 \in \mathbb{R}, x_2 > 0\}$ . The map  $\phi_I$  acts on  $\mathbb{C}^+$  as

$$\phi_I(z) = -\frac{1}{z} = -\frac{\bar{z}}{|z|^2}. \quad (1.25)$$

Thinking of fractional linear transformations, we see that  $\phi_T, \phi_D$  and  $\phi_I$  can be represented as the action of matrices of the form,

$$\gamma \equiv \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \det \gamma = ad - cb > 0, \quad (1.26)$$

on  $\mathbb{C}^+$  given by

$$\gamma z = (az + b)(cz + d)^{-1}. \quad (1.27)$$

We have  $\phi_T, \phi_D$  and  $\phi_I$  implemented by

$$\gamma_T = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix},$$

$$\gamma_D = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix},$$

$$\gamma_I = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

We define  $SL(2, \mathbb{R}) \equiv \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid a, b, c, d \in \mathbb{R}, ad - cb = 1 \right\}$ . Each  $\gamma \in SL(2, \mathbb{R})$  is an isometry of  $\mathbb{H}^2$ . To see this, we compute (1.21) with  $z' = \gamma z, \gamma \in SL(2, \mathbb{R})$ :

$$ds'^2 = \frac{dz' d\bar{z}'}{|\operatorname{Im} z'|^2}.$$

A simple calculation leads to

$$dz' = (cz + d)^{-2} dz, \quad (1.28)$$

so  $|dz'|^2 = |cz + d|^{-4} |dz|^2$ . The coefficient  $|cz + d|^{-2}$  is called the *conformal factor* for  $\gamma$ . Finally, note that

$$\operatorname{Im} z' = \frac{\operatorname{Im} z}{|cz + d|^2}, \quad \operatorname{Re} z' = \frac{ac|z|^2 + (\operatorname{Re} z)(bc + ad) + bd}{|cz + d|^2},$$

so by (1.28) and (1.29), we get

$$ds'^2 = \frac{|dz|^2}{|cz + d|^4} \frac{|cz + d|^4}{|\operatorname{Im} z|^2} = \frac{|dz|^2}{|\operatorname{Im} z|^2} = ds^2.$$

Since  $\gamma$  and  $-\gamma$  implement the same isometry, we see that  $SL(2, \mathbb{R})/\{I, -I\} \equiv PSL(2, \mathbb{R})$  acts as a group of isometries on  $\mathbb{H}^2$ . Below, we will show that this is precisely  $\operatorname{Isom} \mathbb{H}^2$ ,

$$\operatorname{Isom} \mathbb{H}^2 \cong PSL(2, \mathbb{R}), \quad (1.29)$$

which is a real, 3-dimensional Lie group.

**H<sup>3</sup>.** We identify  $x \in \mathbb{H}^3$  with the family of special quaternions of the form

$$z \equiv x_1 + ix_2 + jx_3, \quad (1.30)$$

where  $x = (x_1, x_2, x_3)$  and  $\{i, j, k, 1\}$  span the quaternions. For  $\gamma \equiv \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{C}) = \left\{ \begin{pmatrix} e & f \\ g & h \end{pmatrix} \mid e, f, g, h \in \mathbb{C} \text{ and } eh - fg = 1 \right\}$ , we have an action on such quaternions by

$$z \rightarrow \gamma z = (az + b)(cz + d)^{-1}.$$

Such an action maps  $\mathbb{H}^3$  to itself (identified as in (1.30)) and can be shown to preserve



a real 6-dimensional Lie group.

For the general case, it is convenient to consider  $\mathbb{H}^n$  in the hyperboloid model. In this model, the action of  $\text{Isom } \mathbb{H}^n$  is linear and hence we can easily use group theory to classify  $\text{Isom } \mathbb{H}^n$ . Before we state the main theorem, there is yet another, classical description of  $\text{Isom } \mathbb{H}^n$  which we discuss. This picture provides a strong geometric interpretation of the hyperbolic isometries. We consider the family of transformations of  $\mathbb{R}^{n-1}$  given by reflections in spheres  $S(a, r)$  centered at  $a \in \mathbb{R}^{n-1}$  and of radius  $r$ , and in hyperplanes  $P(b, t)$ , passing through  $b \in \mathbb{R}^{n-1}$  with normal  $t \in \mathbb{R}^{n-1}$ . The latter transformations preserve  $\mathbb{R}^{n-1}$  but not the former. The point  $a$  gets mapped to the point at infinity. If we compactify  $\mathbb{R}^{n-1}$  by adding  $\{\infty\}$ , we obtain  $\hat{\mathbb{R}}^{n-1}$  on which these reflections act in a continuous, invertible manner.

### DEFINITION 1.7

The general Mobius group of  $\mathbb{R}^{n-1}$ ,  $GM(\hat{\mathbb{R}}^{n-1})$ , is the group of transformations generated by reflections in spheres  $S(a, r)$  and planes  $P(b, t)$  in  $\hat{\mathbb{R}}^{n-1}$ . The Mobius group  $M(\hat{\mathbb{R}}^{n-1})$  on  $\hat{\mathbb{R}}^{n-1}$  is the subgroup of  $GM(\hat{\mathbb{R}}^{n-1})$  consisting of orientation-preserving Mobius transformations.

We now consider the map  $\mathbb{R}^{n-1} \hookrightarrow \mathbb{R}^n$  by the injection

$$i: x \in \mathbb{R}^{n-1} \rightarrow \tilde{x} \equiv (x_1, \dots, x_{n-1}, 0),$$

which maps  $\mathbb{R}^{n-1}$  into the  $x_n = 0$  hyperplane. Using this map  $i$ , we can lift any  $\gamma \in GM(\hat{\mathbb{R}}^{n-1})$  to a map on  $\hat{\mathbb{R}}^n$  via its Poincaré extension  $\tilde{\gamma}$ . If  $\gamma$  is a reflection in  $S(a, r)$ , then  $\tilde{\gamma}$  is the reflection in  $S(\tilde{a}, r)$ ; similarly for  $P(b, t)$ . This gives a mapping of  $GM(\hat{\mathbb{R}}^{n-1})$  into  $GM(\mathbb{R}^n)$ . An important fact about Poincaré extensions is (see [B] for a proof)

### PROPOSITION 1.8

The Poincaré extension  $\tilde{\gamma}$  of  $\gamma \in GM(\hat{\mathbb{R}}^{n-1})$  separately preserves  $\mathbb{R}_\pm^n = \{x \in \mathbb{R}^n \mid x_n \gtrless 0\}$ . Furthermore, if  $\beta \in GM(\mathbb{R}^n)$  preserves separately  $\mathbb{R}_\pm^n$  then  $\beta = \tilde{\gamma}$  for some  $\gamma \in GM(\hat{\mathbb{R}}^{n-1})$ .

Consequently, the set of Mobius transformations preserving  $\mathbb{R}_+^n$  can be identified with  $GM(\hat{\mathbb{R}}^{n-1})$ . This connection is deeper, as we will now discuss.

Let  $m \in GM(\hat{\mathbb{R}}^{n-1})$  and let  $\tilde{m}$  be its Poincaré extension. The transformation  $m$  is a finite product of inversions in spheres  $S(a, r)$  and planes  $P(b, t)$ . For  $m = S(a, r)$  (i.e. the reflection in this sphere), the Poincaré extension  $\tilde{m}$  is  $S(\tilde{a}, r)$  where the center  $\tilde{a} = (a_1, \dots, a_{n-1}, 0)$ . From the explicit formula

$$\tilde{m}(x) = \tilde{a} + r^2 \left( \frac{x - \tilde{a}}{\|x - \tilde{a}\|_E^2} \right),$$

$x \in \mathbb{R}_+^n$ , we compute for  $x, y \in \mathbb{R}_+^n, x \neq y$ ,

$$\frac{\|\tilde{m}(x) - \tilde{m}(y)\|_E}{\|x - y\|_E} = \frac{r^2}{\|x - \tilde{a}\|_E \|y - \tilde{a}\|_E} \quad (1.31)$$

and

$$\tilde{m}(x)_n = \frac{r^2 x_n}{\|x - \tilde{a}\|_E^2} \quad (1.32)$$

Consequently, from (1.31)–(1.32) we obtain

$$\frac{\|\tilde{x}(x) - \tilde{m}(y)\|_E^2}{\tilde{m}(x)_n \tilde{m}(y)_n} = \frac{\|x - y\|_E^2}{x_n y_n}. \quad (1.33)$$

For a reflection in a plane,  $P(b, t) \in GM(\hat{\mathbb{R}}^{n-1})$ , its Poincaré extension  $\tilde{m} = P(\tilde{b}, \tilde{t})$ , i.e.  $\tilde{m}$  is a reflection in  $\mathbb{R}^n$  in a plane perpendicular to the plane  $x_n = 0$ . The reflection  $\tilde{m} = P(\tilde{b}, \tilde{t})$  is

$$\tilde{m}(x) = x - 2[(x \cdot \tilde{t}) - b \cdot t] \tilde{t} / \|t\|_E^{-2},$$

from which we see that  $\tilde{m}(x)_n = x_n$  and so

$$\|\tilde{m}(x) - \tilde{m}(y)\|_E = \|x - y\|_E^2, \quad (1.34)$$

so again relation (1.33) holds. So the metric for  $\mathbb{H}^n$  is preserved under any Poincaré extension  $\tilde{m}$  for  $m \in GM(\hat{\mathbb{R}}^{n-1})$ . Combining these observations and Proposition 1.8, we find that  $M(\hat{\mathbb{R}}^{n-1}) < \text{Isom } \mathbb{H}^n$ . We want to prove that these two groups are in fact isomorphic. From (1.33)–(1.34), we see that  $\tilde{m}$  acts conformally on  $\mathbb{R}_+^n$ , i.e. infinitesimally, we have

$$ds'^2 = \frac{d\tilde{m}(x)^2}{\tilde{m}(x)_n^2} = \frac{dx^2}{x_n^2} = ds^2.$$

Now let  $\gamma \in \text{Isom } \mathbb{H}^n$ . The action of  $\gamma$  on  $\mathbb{R}_+^n$  is also conformal, i.e. there exists a positive function  $\lambda(x; \gamma) > 0$  such that

$$d\gamma(x)^2 = \lambda(x; \gamma) dx^2 \quad (1.35)$$

where, explicitly,  $\lambda(x; \gamma) = ((\gamma x)_n / x_n)^2$ . Conformal transformations of an open subset of  $\mathbb{R}^n$  can be completely classified.

**Theorem 1.9 (Liouville's Theorem).** *Let  $f$  be a conformal mapping of an open set  $U \subset \mathbb{R}^n$ ,  $n \geq 3$ , to itself. Then  $f$  is the restriction to  $U$  of an element  $m \in GM(\mathbb{R}^n)$ . We will not discuss the proof of this theorem but refer to a [Ca].*

**COROLLARY 1.10.**

$$M(\hat{\mathbb{R}}^{n-1}) \cong \text{Isom } \mathbb{H}^n, \quad n \geq 3.$$

*Proof.* By (1.35), each  $\gamma \in \text{Isom } \mathbb{H}^n$  is a conformal map of  $\mathbb{R}_+^n \subset \mathbb{R}^n$ . By Liouville's Theorem,  $\gamma$  is the restriction to  $\mathbb{R}_+^n$  of some  $\tilde{m} \in GM(\mathbb{R}^n)$ . Such elements are precisely Poincaré extensions of  $m \in GM(\hat{\mathbb{R}}^{n-1})$ . Consequently, upon restricting to orientation preserving elements, we find that  $M(\hat{\mathbb{R}}^{n-1}) \cong \text{Isom } \mathbb{H}^n$ .  $\square$

We can now prove our main result on the identification of  $\text{Isom } \mathbb{H}^n$  as a classical group.

**Theorem 1.11.** *Isom  $\mathbb{H}^n \cong O^+(n, 1)$ , where  $O^+(n, 1)$  is the connected identity component of the group of real, invertible linear transformations  $K$  such that  $KQK^T = Q$ , with  $Q = \left( \begin{array}{c|c} I_n & 0 \\ \hline 0 & -1 \end{array} \right)$ , and  $K_{n+1, n+1} > 0$ .*

*Proof.* We proceed in two steps and first give the proof for  $n \geq 3$ .

1. Consider the map  $H: \mathbb{R}^n \rightarrow \mathbb{R}^n$  defined by

$$H(x) \equiv (\|x\|_E^2 + 2x_n + 1)^{-1} (2x_1, \dots, 2x_{n-1}, \|x\|_E^2 - 1).$$

One can check that  $H: \mathbb{H}^n \rightarrow B_n$  is an isometric diffeomorphism between  $\mathbb{H}^n$  with the hyperbolic metric (1.2) and  $B_n$  with metric (1.4). By means of this map  $H$ , the group  $GM(\widehat{\mathbb{R}}^{n-1})$  is isomorphic with the subgroup of  $GM(\widehat{\mathbb{R}}^n)$  which preserves  $B_n$ , and similarly for their orientation-preserving subgroups. This latter group is denoted  $GM(B_n)$ , with subgroup  $M(B_n)$ . Then,  $M(B_n) \cong \text{Isom } \mathbb{H}^n$  by these remarks and Corollary 1.10.

2. We show that  $M(B_n) \cong O^+(n, 1)$ , acting as the isometry group of  $Q_n^+$ . Recall that  $F: Q_n^+ \rightarrow B_n$  is the stereographic projection given in (1.12). Since  $O^+(n, 1) < \text{Isom } Q_n^+$ , if  $\gamma \in O^+(n, 1)$  then  $F \circ \gamma \circ F^{-1} \in M(B_n)$ . This mapping is, in fact, a bijection. Any  $\phi \in GM(B_n)$  is of the form  $F \circ \gamma_\phi \circ F^{-1}$  for some  $\gamma_\phi \in O(n, 1)$ , and similarly for  $M(B_n)$  and  $O^+(n, 1)$ . This proves that

$$\text{Isom } \mathbb{H}^n \cong M(B_n) \cong O^+(n, 1) \cong \text{Isom } Q_n^+.$$

We now sketch the proof that the mapping is a bijection. A straight-forward calculation shows that

$$(F \circ \gamma \circ F^{-1})(x) = \frac{(1 + \|x\|_E^2)\gamma_{n+1, j} + 2(x_1\gamma_{1, j} + \dots + x_n\gamma_{n, j})}{\|x\|_E^2(\gamma_{n+1, n+1} - 1) + 2(x_1\gamma_{1, n+1} + \dots + x_n\gamma_{n, n+1}) + (1 + \gamma_{n+1, n+1})}, \quad (1.36)$$

where  $[\gamma_{i, j}] \equiv \gamma \in O(n, 1)$ . Since any  $\phi \in GM(B_n)$  is the composition of an orthogonal transformation and a reflection  $m$  in a sphere orthogonal to  $S^n$ , it suffices to check that each such transformation is of the form  $F \circ \gamma \circ F^{-1}$ . As for orthogonal transformations, if

$$\gamma_A = \left( \begin{array}{c|c} A & 0 \\ \hline 0 & 1 \end{array} \right), \quad A \in O(n),$$

then (1.36) shows that  $(F \circ \gamma_A \circ F^{-1})(x) = Ax$ . Next, it suffices to consider a reflection  $\phi$  in the sphere centered at  $x(t) \equiv (0, 0, \dots, c(t)) \in \mathbb{R}^n$ , with  $c(t) = (\cosh t)(\sinh t)^{-1}$ , and radius  $r = (\sinh t)^{-1}$ ,  $t \in \mathbb{R}$ . The matrix

$$\gamma_\phi \equiv \left( \begin{array}{c|cc} I_{n-1} & & 0 \\ \hline & \cosh 2t & \sinh 2t \\ 0 & \sinh 2t & \cosh 2t \end{array} \right),$$

is easily checked to be in  $O(n, 1)$ . Another calculation of  $F \circ \gamma_\phi \circ F^{-1}$ , as above, shows that this map on  $B_r$  is precisely the reflection in the sphere  $S(x(t), r(t))$ . Since  $F \circ \gamma_\phi \circ F^{-1}$  preserves orientation, the result follows by restriction to  $O^+(n, 1)$ .

3. For the case  $n = 2$ , we have already shown that any element of  $PSL(2, \mathbb{R})$  acts as an isometry of  $\mathbb{H}^2$ . Conversely, suppose that  $\gamma$  is a hyperbolic isometry. For any  $z \in \mathbb{H}^2$ , let  $w = \gamma z$ . For any unit tangent vector  $v$  at  $z$ , let  $\gamma_* v$  be the image tangent vector at  $w$ . We must construct a fractional linear transformation taking  $(z, v)$  to  $(w, \gamma_* v)$ . Since one can map  $z$  to  $w$  by a composition of a rotation and a dilation, there exists a fractional linear transformation mapping  $z$  to  $w$ . This map, however, is not unique since it is determined by two real parameters whereas a general fractional linear transformation is determined by three parameters. The third parameter is fixed by requiring that the unit tangent vector  $v$  at  $z$  be mapped to the unit tangent vector  $\gamma_* v$  at  $w$ . This shows that  $\gamma$  is an element of  $SL(2, \mathbb{R})$ . This completes the proof.  $\square$

Note that for  $n = 2$ , Liouville's Theorem no longer holds: any analytic map of  $\mathbb{R}_+^2$  to itself with  $f' \neq 0$  acts conformally.

#### 1.4 Classification of elements

We classify the elements of  $\text{Isom } \mathbb{H}^n$  according to their fixed points. Let us recall that each  $\gamma \in \text{Isom } \mathbb{H}^n$  arises from an element in  $GM(\hat{\mathbb{R}}^{n-1})$  and hence restricts to a diffeomorphism of  $\partial_\infty \mathbb{H}^n$ . We will use this restriction often but continue to use the symbol  $\gamma$  for  $\gamma|_{\partial_\infty \mathbb{H}^n}$ .

##### DEFINITION 1.12

Let  $\gamma \in \text{Isom } \mathbb{H}^n$ ,  $\gamma \neq id$ .

- (1)  $\gamma$  is called hyperbolic if  $\gamma$  fixes precisely two points on  $\partial_\infty \mathbb{H}^n$ ;
- (2)  $\gamma$  is called parabolic if  $\gamma$  fixes precisely one point on  $\partial_\infty \mathbb{H}^n$ ;
- (3)  $\gamma$  is called elliptic if  $\gamma$  fixes a point of  $\mathbb{H}^n$ .

*Example 1.13.* Consider  $n = 2$  and identify  $\mathbb{H}^2$  with  $\mathbb{C}^+$  as in §1.3. We know that  $\text{Isom } \mathbb{H}^2 \cong PSL(2, \mathbb{R})$ .

$\mathbf{K} \equiv \mathbf{SO}(2)$ , a maximal compact subgroup, consists of elements of the form 
$$\begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}. \quad (1.25)$$
 These elements are all elliptic. For  $\theta = \pi/2$ , we obtain the inversion

$$\gamma_I z = -\frac{1}{z}; \quad \gamma_I = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

This map fixes  $z_0 = +i$ .

$\mathbf{A} \equiv \left\{ \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix} \mid \lambda > 0 \right\}$ . These are the dilations (1.23),  $\gamma_D z = \lambda^2 z$ , and they fix two points on  $\partial_\infty \mathbb{H}^2 = \hat{\mathbb{R}}$ , namely,  $\{0, \infty\}$  and so are hyperbolic.

$N = \left\{ \begin{pmatrix} 1 & \mathbf{b} \\ 0 & 1 \end{pmatrix} \mid \mathbf{b} \in \mathbb{R} \right\}$ . This is the translation subgroup (1.22),

$$\gamma_T z = (x + b) + iy,$$

and fixes  $\{\infty\}$ . Translations are parabolic elements.

Let us note that this is, in fact, the Iwasawa decomposition of  $SL(2, \mathbb{R}) = KAN$ . Any hyperbolic element is conjugate to an element in  $A$ , any parabolic element is conjugate to one in  $N$ , etc. In  $\mathbb{H}^2$ , we can also classify elements according to their trace:

$$\gamma \text{ hyperbolic} \Leftrightarrow |\text{Tr } \gamma| > 2,$$

$$\gamma \text{ parabolic} \Leftrightarrow |\text{Tr } \gamma| = 2,$$

$$\gamma \text{ elliptic} \Leftrightarrow |\text{Tr } \gamma| < 2.$$

In the general case, fixed points of parabolic subgroups will play a special role.

#### DEFINITION 1.14

A fixed point of a parabolic subgroup of  $\text{Isom } \mathbb{H}^n$  is called a cusp.

#### 1.5 Discrete subgroups and the limit set

Theorem 1.11 identifies  $\text{Isom } \mathbb{H}^n$  with a real,  $[n(n+1)]/2$  dimensional Lie group. Hence,  $\text{Isom } \mathbb{H}^n$  is naturally a topological group with the metric topology on  $O^+(n, 1)$ .

#### DEFINITION 1.15

A subgroup  $\Gamma < \text{Isom } \mathbb{H}^n$  is discrete if the identity  $I \equiv id$  is isolated from  $\Gamma \setminus \{I\}$  in the group topology.

It follows by the continuity of the group operations that if  $\Gamma$  is discrete, then the relative topology on  $\Gamma$  is the discrete topology, i.e.  $\{\gamma\}$ ,  $\gamma \in \Gamma$ , is open in this topology. The converse is also true. Moreover, since  $O^+(n, 1)$  is a matrix group, its topology is compatible with the norm topology. From this we can deduce that if  $\Gamma < \text{Isom } \mathbb{H}^n$  is discrete, then it is countable.

We are interested in discrete subgroups of  $\text{Isom } \mathbb{H}^n$  because they have a particularly nice action on  $\mathbb{H}^n$ .

#### DEFINITION 1.16

A subgroup  $\Gamma < \text{Isom } \mathbb{H}^n$  acts discontinuously on  $\mathbb{H}^n$  if for any compact subset  $K \subset \mathbb{H}^n$ ,  $\gamma K \cap K \neq \emptyset$  for at most finitely-many  $\gamma \in \Gamma$ . (Here,  $\gamma K = \{\gamma x \mid x \in K\}$ ).

We can reformulate this in terms of  $\Gamma$ -orbits. For  $x \in \mathbb{H}^n$ ,  $\Gamma_x = \{\gamma x \mid \gamma \in \Gamma\}$  is called the  $\Gamma$ -orbit of  $x$ . If  $\Gamma$  acts discontinuously on  $\mathbb{H}^n$ , then each compact subset  $K \subset \mathbb{H}^n$  contains only finitely-many points of  $\Gamma_x$ , for each  $x \in K$ . The most important aspect of discreteness of  $\Gamma$  is its influence on the action of  $\Gamma$  on  $\mathbb{H}^n$ .

#### PROPOSITION 1.17

*Let  $\Gamma < \text{Isom } \mathbb{H}^n$  be discrete. Then  $\Gamma$  acts discontinuously on  $\mathbb{H}^n$ . The converse is also true.*

A proof of this result for  $n = 3$  can be found in [B]. (The proof of the converse is easy). Discrete subgroups of  $\text{Isom } \mathbb{H}^2$  are called Fuchsian groups and discrete subgroups of  $\text{Isom } \mathbb{H}^3$  are called Kleinian.

Another consequence of discreteness is the characterization of the accumulation points of  $\Gamma$ -orbits. Since  $\Gamma$  acts discontinuously, these accumulation points can only lie in  $\partial_\infty \mathbb{H}^n = \hat{\mathbb{R}}^{n-1}$ .

### DEFINITION 1.18

Let  $\Gamma < \text{Isom } \mathbb{H}^n$  be discrete. The limit set of  $\Gamma$ ,  $\Lambda(\Gamma)$ , is the subset of  $\partial_\infty \mathbb{H}^n$  consisting of all the accumulation points of all  $\Gamma$ -orbits.

Since  $\Gamma$  acts naturally on  $\partial_\infty \mathbb{H}^n$ , we can characterize  $\Lambda(\Gamma)$  as

- (1)  $\Lambda(\Gamma)$  is the smallest, closed  $\Gamma$ -invariant subset of  $\partial_\infty \mathbb{H}^n$ ;
- (2)  $\Lambda(\Gamma)$  contains all parabolic and hyperbolic fixed points of subgroups of  $\Gamma$ .

The complement of  $\Lambda(\Gamma)$  in  $\partial_\infty \mathbb{H}^n$  is called the **domain of discontinuity** and denoted by  $\Omega(\Gamma) \equiv \partial_\infty \mathbb{H}^n \setminus \Lambda(\Gamma)$ . The domain of discontinuity is the largest open subset of  $\partial_\infty \mathbb{H}^n$  on which  $\Gamma$  acts discontinuously.

We will not discuss  $\Lambda(\Gamma)$  further, but mention some additional facts.

- (1) For a Kleinian group  $\Gamma$  which is geometrically finite, (see Definition 1.21), the classical exponent of convergence  $\delta(\Gamma)$  (which lies in  $[0, 2]$ ) for the Poincaré series (see section 4) is equal to the Hausdorff dimension of the limit set.
- (2)  $\Lambda(\Gamma)$  very often has a Cantor-like structure, i.e. it is an uncountable, perfect set.
- (3) For Kleinian groups  $\Gamma$  as in (1), if  $\Delta_\Gamma$  is the Laplacian on  $\mathcal{M} \equiv \mathbb{H}^3/\Gamma$  (see section 1.7) then  $\inf \sigma(\Delta_\Gamma) = \delta(\Gamma)(2 - \delta(\Gamma))$ , where  $\sigma(\Delta_\Gamma)$  is the spectrum of the Laplacian.

Results (1) and (2) are quite deep and due to Patterson [Pal] and Sullivan [S]. Patterson has also proved that  $\Lambda(\Gamma)$  supports a  $\Gamma$ -invariant measure (see also [S] for similar results).

### 1.6 Fundamental domains

Let  $\Gamma < \text{Isom } \mathbb{H}^n$  be a discrete subgroup. We know that any compact  $K \subset \mathbb{H}^n$  contains only finitely-many points of each  $\Gamma$ -orbit. We want to cut down the size of  $K$  to obtain a subset of  $\mathbb{H}^n$  containing exactly one representative of each  $\Gamma$ -orbit.

### DEFINITION 1.19

An open, connected subset  $F \subset \mathbb{H}^n$  is a **fundamental domain** for a discrete subgroup  $\Gamma < \text{Isom } \mathbb{H}^n$  if

- (i) no two points of  $F$  are  $\Gamma$ -equivalent;
- (ii) if  $z_1, z_2 \in \bar{F}$  are such that  $z_1 = \gamma z_2$ ,  $\gamma \in \Gamma$  then  $z_1, z_2 \in \partial F$ ;
- (iii)  $\bigcup_{\gamma \in \Gamma} \gamma \bar{F} = \mathbb{H}^n$  (closure is in the hyperbolic metric).

If  $F$  is a fundamental domain for a discrete group  $\Gamma$ , so is any set  $\gamma F$ ,  $\gamma \in \Gamma$ , and we say that  $\{\gamma F | \gamma \in \Gamma\}$  forms a tiling of  $\mathbb{H}^n$ . Note that by construction,  $F$  contains exactly one point from each  $\Gamma$ -orbit.

Examples 1.20.  $\mathbb{H}^2$

(i) Translations:  $\gamma_T \equiv \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$  and  $\Gamma = \langle \gamma_T \rangle$  (cyclic subgroup generated by  $\gamma_T$ ). A fundamental domain is the strip

$$F = \{(x_1, x_2) | x_2 > 0, -\frac{1}{2} < x_1 < \frac{1}{2}\}.$$

(ii) Dilations:  $\gamma_D = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}$ ,  $\Gamma = \langle \gamma_D \rangle$ . The annular region

$$F \equiv \{(x_1, x_2) | x_2 > 0, 1 < \sqrt{x_1^2 + x_2^2} < \lambda^2\},$$

is a fundamental domain for  $\Gamma$ .

(iii)  $\Gamma = PSL(2, \mathbb{Z})$ , the modular group.  $\Gamma$  is generated by the parabolic motion  $\gamma_T$  as in (i) and the inversion in the unit semi-circle

$$\gamma_I = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

A fundamental domain for  $\gamma_I$  is

$$F_I = \{(x_1, x_2) | 1 < x_1^2 + x_2^2\}.$$

Combining this  $F_I$  and the strip in (i) one sees that a fundamental domain for  $\Gamma$  is

$$F = \{(x_1, x_2) | 1 < x_1^2 + x_2^2 \text{ and } -\frac{1}{2} < x_1 < \frac{1}{2}\}.$$

Note that  $\{\infty\}$  is a fixed point for  $\gamma_T$ . If we reflect  $F$  in the unit semi-circle, we map  $\{\infty\}$  to 0, and we see that  $F$  has a cusp at 0 (the cusp corresponding to the parabolic fixed point for the subgroup of  $\Gamma$  generated by  $\gamma_I \gamma_T \gamma_I$ ).

Given a fundamental domain for  $\Gamma$ , we can compute its volume using the volume form given by

$$dV = x_n^{-n} dx_1 \dots dx_n. \quad (1.37)$$

For the examples above,  $\text{Vol}(F) = \infty$  for  $\langle \gamma_T \rangle$  and  $\langle \gamma_D \rangle$  whereas  $\text{Vol}(F) < \infty$  for  $PSL(2, \mathbb{Z})$ . This can be seen as follows. For any  $a > 1$ .

$$\int_{-1/2}^{1/2} \int_a^\infty \frac{dx_1 dx_2}{x_2^2} = a^{-1} < \infty,$$

so that

$$\int_F dV = a^{-1} + (\text{finite piece}).$$

In general,  $\text{Vol}(F) < \infty$  if  $\partial_\infty \mathbb{H}^n \cap (F)^{\text{cl}}$  is discrete, where cl denotes the Euclidean closure.

We will restrict the class of discrete groups that we study here by the following condition.

$\Gamma \subset \mathbb{H}^n$  bounded by finitely many geodesic hypersurfaces.

Recall from the discussion of geodesics, that a geodesic hypersurface is a hemi-sphere with center on  $\partial_\infty \mathbb{H}^n$  (so it includes hyperplanes perpendicular to  $x_n = 0$ ). We develop a rather complete theory for geometrically finite, discrete subgroups. The theory for "geometrically infinite" discrete subgroups still remains to be developed. Some interesting situations are studied by Epstein in [Ep] (see Chapter 5).

We conclude this section with a description of the canonical construction of a fundamental domain called the *Dirichlet Polygon*. Let us assume that  $\Gamma$  is discrete and contains no elliptic elements. Let  $w \in \mathbb{H}^n$ . We define sets for each  $\gamma \in \Gamma$ :

$$L_\gamma(w) = \{z \in \mathbb{H}^n \mid d_h(z, w) = d_h(z, \gamma w)\}.$$

This hypersurface divides  $\mathbb{H}^n$  into two disjoint sets,

$$H_\gamma(w) \equiv \{z \in \mathbb{H}^n \mid d_h(z, w) < d_h(z, \gamma w)\},$$

which contains  $w$ , and its complement. One can check that  $\mathbb{H}^n \setminus \{H_\gamma(w) \cup L_\gamma(w) \mid \gamma \in \Gamma, \gamma \neq \text{id}\}$  is a fundamental domain for  $\Gamma$ .  $H_{\gamma^{-1}}(\gamma w)$  and that  $L_\gamma(w)$  is, in fact, a geodesic hypersurface.

#### DEFINITION 1.22

The Dirichlet Polygon for  $\Gamma$  based at  $w \in \mathbb{H}^n$  is the set

$$D_\Gamma(w) \equiv \bigcap_{\substack{\gamma \in \Gamma \\ \gamma \neq \text{id}}} H_\gamma(w).$$

We see that  $D_\Gamma(w)$  is the set of  $z \in \mathbb{H}^n$  closer to  $w$  than any point  $\gamma w, \gamma \in \Gamma$ . The Dirichlet polygons satisfy several nice properties relative to  $\Gamma$ . For example, if  $\gamma \in \Gamma$ , then

$$\gamma D_\Gamma(w) = D_\Gamma(\gamma w),$$

which follows from the observation that

$$\gamma H_{\tilde{\gamma}}(w) = H_{\gamma \tilde{\gamma} \gamma^{-1}}(\gamma w),$$

for  $\gamma, \tilde{\gamma} \in \Gamma$ . The main fact about the Dirichlet polygon is the following, which we mention without proof (see [B]).

**Theorem 1.23.** *The Dirichlet polygon  $D_\Gamma(w)$  is a convex fundamental domain for  $\Gamma$ . Furthermore, for any  $K \subset \mathbb{H}^n$  compact, the set of  $\gamma \in \Gamma$  such that  $\gamma D_\Gamma(w)^{\text{cl}} \cap K \neq \emptyset$  is finite.*

We remark that even if  $\Gamma$  is geometrically finite, it may happen that  $D_\Gamma(w)$  is not bounded by finitely many geodesically hypersurfaces.

#### 1.7 Uniformization theorem for hyperbolic manifolds

We now turn to the general theory of hyperbolic manifolds.



## DEFINITION 1.24

An  $n$ -dimensional manifold  $\mathcal{M}$  is an  $n$ -dimensional hyperbolic manifold if  $\mathcal{M}$  is a complete, connected real Riemannian manifold with all sectional curvatures equal to  $-1$ .

Hyperbolic manifolds that are not simply connected can be constructed from  $\mathbb{H}^n$  and discrete subgroups  $\Gamma < \text{Isom } \mathbb{H}^n$  with no elliptic elements as follows. By § 1.6, we can choose a fundamental domain  $F_\Gamma \subset \mathbb{H}^n$ . We identify points  $z_1, z_2 \in \partial F_\Gamma$  if  $z_2 = \gamma z_1$  for  $\gamma \in \Gamma$ . We equip this set with the metric induced by the hyperbolic metric (1.2) on  $\mathbb{H}^n$ . This results in an  $n$ -dimensional real Riemannian manifold  $\mathcal{M}_\Gamma$ . The manifold  $\mathcal{M}_\Gamma$  is geodesically complete since any geodesic can be infinitely extended. Any geodesic starting at  $p \in F_\Gamma$  can be extended across  $\partial F_\Gamma$  (since  $\mathbb{H}^n$  is complete) and by the identification of points on  $\partial F_\Gamma$ , the geodesic will be closed or it will be mapped to the intersection of  $F_\Gamma$  with another geodesic of  $\mathbb{H}^n$ . The Hopf-Rinow Theorem allows us to conclude that  $\mathcal{M}_\Gamma$  is complete. Since the metric is locally (1.2), it is clear that all sectional curvatures are  $-1$ . It is not too difficult to see that the fundamental group  $\pi_1(\mathcal{M}_\Gamma)$  is isomorphic to  $\Gamma$ . In fact, the homotopy classes of closed geodesics of  $\mathcal{M}_\Gamma$  are in one-to-one correspondence with the conjugacy classes of  $\Gamma$ . We refer to  $\mathcal{M}_\Gamma$  as the quotient manifold  $\mathcal{M}_\Gamma = \mathbb{H}^n/\Gamma$ .

This construction of hyperbolic manifolds is quite general.

**Theorem 1.25** *Let  $\mathcal{M}$  be an  $n$ -dimensional hyperbolic manifold. Then there exists a discrete subgroup  $\Gamma < \text{Isom } \mathbb{H}^n$  with no elliptic elements such that  $\mathcal{M} = \mathbb{H}^n/\Gamma$  and  $\pi_1(\mathcal{M}) \cong \Gamma$ .*

*Sketch of the Proof.* Let  $\tilde{\mathcal{M}}$  be the universal cover of  $\mathcal{M}$ , i.e.  $\tilde{\mathcal{M}}$  is a complete,  $n$ -dimensional simply connected Riemannian manifold with constant sectional curvatures equal to  $-1$ . Because of the negative curvature and simply connectivity, the exponential map is a diffeomorphism of  $\tilde{\mathcal{M}}$  with  $\mathbb{R}^n$ , identified as  $T_p(\tilde{\mathcal{M}})$ , by the theorem of Cartan and Hadamard. Similarly, the exponential map is a diffeomorphism of  $\mathbb{H}^n$  with  $\mathbb{R}^n$ . If we choose a linear isometry  $i: \mathbb{R}^n \rightarrow \mathbb{R}^n$ , then the map  $f \equiv \exp_{\tilde{\mathcal{M}}, p} \circ i \circ \exp_{\mathbb{H}^n, p}^{-1}: \mathbb{H}^n \rightarrow \tilde{\mathcal{M}}$  is an isometry of  $\tilde{\mathcal{M}}$  and  $\mathbb{H}^n$  by a theorem of Cartan-Hadamard. By construction, the group  $\Gamma$  of covering transformations are isometries of  $\tilde{\mathcal{M}}$ . These isometries act freely and discontinuously on  $\tilde{\mathcal{M}} \cong \mathbb{H}^n$ , so by Theorem 1.17 they form a discrete subgroup of  $\text{Isom } \mathbb{H}^n$ . To conclude the proof, one identifies  $\tilde{\mathcal{M}}/\Gamma$  with  $\mathcal{M}$  and  $\pi_1(\tilde{\mathcal{M}}/\Gamma) \cong \pi_1(\mathcal{M}) \cong \Gamma$ .

We conclude this discussion by giving a global classification of hyperbolic manifolds which will be useful in the spectral analysis. We will consider compact and non-compact manifolds  $\mathcal{M} = \mathbb{H}^n/\Gamma$ . Compact manifolds are always of finite volume whereas we distinguish between non-compact manifolds of finite and infinite volume.

It is known that if  $\mathcal{M} = \mathbb{H}^n/\Gamma$  is compact, then  $\Gamma$  contains only hyperbolic elements. An example of a non-compact manifold of finite volume is given by  $\mathbb{H}^2/\text{PSL}(2, \mathbb{Z})$  and one of infinite volume is given by  $\mathbb{H}^2/\langle \gamma_D \rangle$ , where  $\gamma_D$  is a dilation.

Elliptic elements give rise to singularities of the quotient  $\mathbb{H}^n/\Gamma$ . Such quotients are not manifolds but are called *orbitals*. Groups  $\Gamma$  with no elliptic elements are called *torsion-free*.

## 2. Spectral preliminaries

This section is devoted to a preliminary spectral analysis of geometrically finite hyperbolic manifolds  $\mathcal{M} = \mathbb{H}^n / \Gamma$ . We will always assume  $\Gamma < \text{Isom } \mathbb{H}^n$  is discrete, contains no elliptic elements, and is geometrically finite. We discuss the general spectral characteristics of compact and of non-compact manifolds. We will refine these results in chapter 3 for non-compact manifolds. There, we study the generalized eigenfunctions and use them to construct the spectral representation of the Laplace–Beltrami operator  $\Delta_\Gamma$  on  $\mathcal{M}$ . The goals of this chapter are to:

- (a) characterize the spectrum of compact manifolds;
- (b) identify the essential spectrum of non-compact manifolds;
- (c) prove the absence of singular continuous spectrum;
- (d) characterize the discrete spectrum.

The last two parts rely heavily on a technique called the Mourre theory. We review the main points of this method in § 2.4. In order to apply this method, need to develop the asymptotic geometry at infinity of non-compact, hyperbolic manifolds.

### 2.1 The Hilbert space $L^2(\mathcal{M}, dV)$ and the Laplace–Beltrami operator

**A. Integration on manifolds.** Let  $(\mathcal{M}, g)$  be a Riemannian manifold with metric  $g$ . Let  $\{(U_i, \phi_i)\}$  be a set of coordinate charts for  $\mathcal{M}$ , i.e.  $U_i \subset \mathcal{M}$  is open,  $\mathcal{M} = \bigcup_i U_i$ , and  $\phi_i: U_i \rightarrow \mathcal{O}_i \subset \mathbb{R}^n$ , where  $\mathcal{O}_i$  is open. If  $\mathcal{M}$  is orientable (as for hyperbolic manifolds) there exists an everywhere positive  $n$ -form, called the volume form,  $dV$ . For hyperbolic manifolds, it is given in local coordinates by (1.37). Suppose that  $(\mathcal{M}, g)$  is a differentiable manifold, which means that the maps  $\phi_i \circ \phi_j^{-1}: \phi_j(U_j \cap U_i) \rightarrow \phi_i(U_j \cap U_i)$  of open subsets of  $\mathbb{R}^n$  are differentiable. If  $f: \mathcal{M} \rightarrow \mathbb{C}$  is a smooth function of compact support, we can define the  $n$ -form  $f dV$ . This can be integrated over  $\mathcal{M}$  as follows. Let  $\{\chi_i\}$  be a partition of unity for  $\mathcal{M}$  subordinate to the open cover  $\{U_i\}$ . We have  $f = \sum_i f \chi_i$  where the sum is finite. We then define the integral

$$\int_{\mathcal{M}} f dV \equiv \sum_i \int_{\phi_i(U_i)} (\chi_i f \circ \phi_i^{-1})(x) dV(x). \quad (2.1)$$

This integral exists and is independent of the partition of unity.

We assume that  $(\mathcal{M}, g)$  has a  $C^\infty$ -differentiable structure. This is true for  $\mathcal{M} = \mathbb{H}^n / \Gamma$  (see the discussion § 2.5). Let  $C_0^\infty(\mathcal{M})$  be the linear vector space of all complex-valued, smooth functions of compact support on  $\mathcal{M}$ . By means of the integral (2.1), we can define a norm on  $C_0^\infty(\mathcal{M})$  by

$$\|f\| \equiv \left[ \sum_i \int_{\phi_i(U_i)} |(\chi_i f \circ \phi_i^{-1})(x)|^2 dV(x) \right]^{1/2}. \quad (2.2)$$

This norm is induced by a naturally associated inner product. When  $\mathcal{M} = \mathbb{R}^n$  with the Euclidean metric, this is just the  $L^2(\mathbb{R}^n)$  norm.

## DEFINITION 2.1

The Hilbert space  $L^2(\mathcal{M}, dV)$  is the completion of  $C_0^\infty(\mathcal{M})$  in the norm (2.2).

We mention another way to think of  $L^2(\mathcal{M}, dV)$  when  $\mathcal{M} = \mathbb{H}^n/\Gamma$  is hyperbolic (or, more generally, a symmetric space).

## DEFINITION 2.2

A function  $f: \mathbb{H}^n \rightarrow \mathbb{C}$  is  $\Gamma$ -automorphic if  $f(\gamma x) = f(x) \forall x \in \mathbb{H}^n, \forall \gamma \in \Gamma$ .

Let  $F_\Gamma$  be a fundamental domain for  $\Gamma$ . Any  $f \in C_0(F_\Gamma)$  extends naturally to a  $\Gamma$ -automorphic function on  $\mathbb{H}^n$ . Using the global coordinates on  $\mathbb{H}^n$ , we can define a norm on  $\Gamma$  automorphic functions by

$$\|f\|^2 \equiv \int_{F_\Gamma} |f(x)|^2 \frac{dx_1 \dots dx_n}{x_n^n}. \quad (2.3)$$

The completion of  $(C_0(F_\Gamma), \|\cdot\|)$  is just the Hilbert space of Definition 2.1. We will often write  $d\mu$  for the hyperbolic volume element occurring in (2.3).

**B. The Laplace–Beltrami operator.** For a smooth Riemannian manifold  $(\mathcal{M}, g)$ , we define a second-order, non-negative elliptic operator by

$$\Delta_g f \equiv -\operatorname{div}(\operatorname{grad} f), \quad (2.4)$$

for  $f \in C_0^\infty(\mathcal{M})$ . In local coordinates, one can write

$$\Delta_g \equiv -\frac{1}{\sqrt{\det g}} \sum_{i,j=1}^n \frac{\partial}{\partial x_i} g^{ij} \sqrt{\det g} \frac{\partial}{\partial x_j}, \quad (2.5)$$

where  $\det g$  is the determinant of the metric  $g$ . This operator is called the *Laplace–Beltrami operator*, or simply, the *Laplacian*, for  $(\mathcal{M}, g)$ . It defines a non-negative symmetric operator on the dense set  $C_0^\infty(\mathcal{M})$  in the Hilbert space  $L^2(\mathcal{M}, \sqrt{\det g} dx_1 \wedge \dots \wedge dx_n)$ . This operator is the main object of our study.

**Examples 2.3.** (1)  $\mathcal{M} = \mathbb{R}^n$  with the Euclidean metric, then

$$\Delta_E \equiv -\sum_{i=1}^n \frac{\partial^2}{\partial x_i^2}.$$

It is well-known that  $\Delta_E$  is self-adjoint on the domain  $H^2(\mathbb{R}^n)$ .

(2)  $\mathcal{M} = \mathbb{H}^n$ . The Laplacian is constructed from the vector fields  $x_n \partial / \partial x_i$ . From (1.2) and (2.5) one easily computes

$$\Delta_{\mathbb{H}^n} \equiv -\left(x_n \frac{\partial}{\partial x_n}\right)^2 + (n-1)x_n \frac{\partial}{\partial x_n} + x_n^2 P, \quad (2.6)$$

with  $P$  the Euclidean Laplacian on  $\mathbb{R}^{n-1}$ ,

$$P \equiv -\sum_{i=1}^{n-1} \frac{\partial^2}{\partial x_i^2}. \quad (2.7)$$

For general hyperbolic manifolds, it follows from Theorem 1.25 that (2.6) represents the Laplacian in local coordinates. Note that the coefficients of  $\Delta_{\mathbb{H}^n}$  degenerate on  $\partial_\infty \mathbb{H}^n$ . The operator (2.6) is self-adjoint (see Theorem 2.4 below).

To study the spectral properties of the Laplacian on  $(\mathcal{M}, g)$ , it is essential that it be self-adjoint. The following theorem of Gaffney [G] suffices.

**Theorem 2.4.** [G] *Let  $(\mathcal{M}, g)$  be a smooth, complete Riemannian manifold without boundary. Then the Laplacian  $\Delta_g \equiv -\operatorname{div}(\operatorname{grad})$  (given locally by (2.5)) is essentially self-adjoint on  $C_0^\infty(\mathcal{M})$ .*

The domain of self-adjointness can be characterized in terms of the Sobolev spaces  $H^s(\mathcal{M})$ .

#### DEFINITION 2.5.

Let  $(\mathcal{M}, g)$  be a smooth Riemannian manifold without boundary and let  $\nabla$  denote the gradient. For  $s \in \mathbb{N}$  and  $g \in C_0^\infty(\mathcal{M})$ , we define, for a multi-index  $l = (l_1, \dots, l_n)$ ,

$|l| = \sum_{i=1}^n l_i$ , a norm

$$\|g\|_s \equiv \sum_{|l| \leq s} \|\nabla^l g\|,$$

where  $\|\cdot\|$  is defined in (2.2). The completion of  $C_0^\infty(\mathcal{M})$  in the norm (2.8) is the Sobolev space  $H^s(\mathcal{M})$ .

#### COROLLARY 2.6.

*Under the assumption of Theorem 2.4,  $\Delta_g$  is self-adjoint on  $L^2(\mathcal{M})$  with domain  $H^2(\mathcal{M})$ .*

For a description of the spectral properties of a self-adjoint operator, we refer the reader to [RS2, 4]. We follow the notation and definition of these texts. In particular, the spectrum of  $A$  is denoted by  $\sigma(A)$ , etc. For  $\mathbb{H}^n$ , let us note that the gradient is given by  $\nabla_i \equiv x_n \frac{\partial}{\partial x_i}$ ,  $i = 1, \dots, n$ . Combining the above comments, we obtain

#### COROLLARY 2.7.

*Let  $\mathcal{M} \equiv \mathbb{H}^n/\Gamma$  be a hyperbolic manifold. The Laplacian on  $\mathcal{M}$ , which we now write as  $\Delta_\Gamma$ , is self-adjoint on  $H^2(\mathcal{M})$  and its spectrum,  $\sigma(\Delta_\Gamma)$ , is a closed subset of  $[0, \infty)$ .*

To conclude this discussion, we mention the Sobolev embedding theorems for smooth manifolds without boundary and for bounded open subsets of manifolds (see [Ta]).

**Theorem 2.8.** *Let  $\mathcal{M}$  be a smooth, compact Riemannian manifold. The imbedding  $H^s(\mathcal{M}) \hookrightarrow H^{s-2}(\mathcal{M})$ ,  $s \geq 2$ , is compact. If  $\Omega \subset \mathcal{M}$  is a bounded open set with  $C^1$ -boundary, then the imbedding  $H^s(\Omega) \hookrightarrow H^{s-2}(\Omega)$ ,  $s \geq 2$ , is compact.*

### 2.2 The spectra of compact hyperbolic manifolds

The general characterization of the spectrum of  $\Delta_\Gamma$  on a compact hyperbolic manifold is rather easy. There are, however, many deep results concerning the relation between

the spectrum and the geometry which we will mention. In this section,  $\mathcal{M} = \mathbb{H}^n/\Gamma$  is a compact hyperbolic manifold. The volume is necessarily finite and all elements in  $\Gamma$  are hyperbolic. As an example, we can take certain groups generated by reflections in finitely-many hemispheres. The main theorem is

**Theorem 2.9.**  $\Delta_\Gamma$  on  $\mathcal{M}$  has only discrete spectrum with eigenvalues  $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \dots$  of finite multiplicity and having  $\infty$  as the only point of accumulation. The eigenfunctions  $\{\phi_k\}$  form an orthonormal basis of  $L^2(\mathcal{M}, dV)$ , so for any  $f \in L^2(\mathcal{M}, dV)$  the series

$$f = \sum_{k=0}^{\infty} \langle \phi_k, f \rangle \phi_k,$$

converges in norm. The Laplacian  $\Delta_\Gamma$  is diagonalizable. If  $\mathcal{E}_j$  is the finite-dimensional eigenspace for eigenvalue  $\lambda_j \in \sigma(\Delta_\Gamma)$  and  $P_j: L^2(\mathcal{M}, dV) \rightarrow \mathcal{E}_j$  is the corresponding spectral projection, then

$$\Delta_\Gamma = \sum_{j=0}^{\infty} \lambda_j P_j.$$

Finally, the eigenfunctions  $\phi_k \in C^\infty(\mathcal{M})$ .

*Proof.* By Corollary 2.7,  $\Delta_\Gamma \geq 0$  and  $D(\Delta_\Gamma) = H^2(\mathcal{M}) = (\Delta_\Gamma + 1)^{-1} L^2(\mathcal{M})$ . The resolvent  $(\Delta_\Gamma + 1)^{-1}$  is a bounded self-adjoint operator. By the Sobolev embedding theorem, Theorem 2.8, the identity map  $i: H^2(\mathcal{M}) \rightarrow L^2(\mathcal{M})$  is compact. Consequently the product  $i \cdot (\Delta_\Gamma + 1)^{-1} \in \mathcal{B}(L^2(\mathcal{M}))$  is a compact, non-negative self-adjoint operator. Standard spectral results (see [RS1]) state that  $\sigma((\Delta_\Gamma + 1)^{-1}) \equiv \{\rho_k\}$  is purely discrete,  $\rho_k \geq 0$ , and with zero the only possible accumulation point of the eigenvalues. If  $\phi_k$  is an eigenfunction, then  $(\Delta_\Gamma + 1)^{-1} \phi_k = \rho_k \phi_k$ . As  $(\Delta_\Gamma + 1)^{-1} \phi_k \in H^2(\mathcal{M})$ , we have  $\phi_k \in H^2(\mathcal{M})$  and

$$\phi_k = (\Delta_\Gamma + 1)[(\Delta_\Gamma + 1)^{-1} \phi_k] = \rho_k(\Delta_\Gamma + 1)\phi_k,$$

i.e. if  $\rho_k \neq 0$ ,

$$\Delta_\Gamma \phi_k = \left( \frac{1 - \rho_k}{\rho_k} \right) \phi_k.$$

Now  $\Delta_\Gamma \geq 0$  so  $\|(\Delta_\Gamma + 1)^{-1}\| \leq 1$  so  $\lambda_k \equiv (1 - \rho_k)\rho_k^{-1}$  is a non-negative eigenvalue of  $\Delta_\Gamma$  with finite-multiplicity. It is easy to see that, in fact,  $\rho_k > 0$ , i.e.  $\ker(\Delta_\Gamma + 1)^{-1} = \emptyset$ . As  $\rho_k = 1$  corresponds to  $\phi_k \equiv \text{constant}$ ,  $0 \in \sigma(\Delta_\Gamma)$ . The eigenvalues  $\{\lambda_k\}$  can accumulate only at infinity. The completeness of the eigenfunctions of  $(\Delta_\Gamma + 1)^{-1}$  imply that for  $\Delta_\Gamma$ . The expansion statements are now standard. Finally, since  $\Delta_\Gamma$  is an elliptic operator on  $\mathcal{M}$  with  $C^\infty$ -coefficients it follows by standard elliptic regularity results (see [RS4]) that  $\psi_k \in C^\infty(\mathcal{M})$ .  $\square$

Although Theorem 2.9 is rather elementary, the nature of the relationship between the spectrum  $\{\lambda_k\}$  and the geometry and topology of  $\mathcal{M}$  is quite subtle and still an area of active research. A large part of the book of Chavel [Ch] is devoted to these questions. We mention a few of the results and problems here.

1.  $n = 2$ . Suppose  $\mathbb{H}^2/\Gamma$  is a compact Riemann surface of genus  $g \geq 2$ . It is known [SWY] that for any  $\varepsilon > 0 \exists \Gamma < \text{PSL}(2, \mathbb{R})$  so that  $\lambda_{2g-3} < \varepsilon$ . So for a surface of genus 2, the first non-zero eigenvalue can be made arbitrarily small. Buser [Bu] proved a lower-bound  $\lambda_{4g-2} \geq \frac{1}{4}$ . In general,  $\lambda_1 \leq 2(g+1)(g-1)^{-1} \leq 6$ .

2. Selberg trace formula. This fundamental result connects the spectrum of  $\Delta_\Gamma$  and the length spectrum of  $\mathcal{M} \equiv \mathbb{H}^n/\Gamma$ , i.e. the length of closed geodesics on  $\mathcal{M}$ . If  $e^{-t\Delta_\Gamma}$ ,  $t \geq 0$ , is the heat operator on  $\mathcal{M}$ , it is trace class. If its distribution kernel is written  $K_\Gamma(x, y, t)$ , we have

$$\begin{aligned} \text{Tr}(e^{-t\Delta_\Gamma}) &= \int_{\mathcal{M}} K_\Gamma(x, x, t) dV(x) \\ &= \text{vol}(\mathcal{M}) + \sum_{n=1}^{\infty} \sum_{\substack{\text{inconjugate} \\ \text{primitive } p \in \Gamma}} \frac{l(p)}{[\cosh l(p^n) - 1]^{1/2}} \\ &\quad \times \int_{\cosh l(p^n)}^{\infty} \frac{K_\Gamma(b) db}{[b - \cosh l(p^n)]^{1/2}}. \end{aligned}$$

Here, an inconjugate element  $p \in \Gamma$  is one which is not conjugate to any power of another element of  $\Gamma$ . Associated with such elements  $p$  are homotopy classes of closed geodesics. We denote by  $l(p)$  the length of the minimal closed geodesic in such a class. For  $n \in \mathbb{N}$ ,  $p^n$  labels the conjugacy classes of  $\Gamma$ . Finally, since  $K_\Gamma$  is a function of the hyperbolic distance  $d_h(x, y)$ , we write  $K_\Gamma(b)$  in the formula.

An application of this formula is the asymptotic behavior of the eigenvalue counting function  $N_\Gamma(\lambda)$  defined as

$$N_\Gamma(\lambda) \equiv \{ \# \lambda_j | \lambda_j \in \sigma(\Delta_\Gamma) \text{ and } \lambda_j \leq \lambda \}.$$

One can prove that as  $\lambda \rightarrow \infty$ ,

$$N_\Gamma(\lambda) \sim \frac{\text{vol}(\mathcal{M})}{\omega_n} \lambda^{n/2},$$

where  $\sim$  means “asymptotic to” and  $\omega_n$  is the Euclidean volume of the unit ball in  $\mathbb{R}^n$ . This behavior is referred to as a “Weyl law”. We refer to the expository article of McKean [McK] for the proofs.

3. We mention that for general compact, smooth Riemannian manifolds without boundary a lot of work has been devoted to the isospectral problem: suppose  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are isospectral, are they isomorphic? We refer to the review article to Perry [Pel] for a discussion. This problem was popularized by an article of M. Kac “Can one hear the shape of a drum?” [K]. It has only recently been proved that there exist a pair of isospectral, non-isomorphic planar domains [GWW].

### 2.3 The spectra of non-compact, geometrically finite hyperbolic manifolds: The main theorem

The goal of the remainder of this chapter and of chapter 3 is to prove the analog of Theorem 2.9 for an arbitrary,  $n$ -dimensional, geometrically finite hyperbolic manifold. In this section, we will discuss the spectrum of  $\Delta_\Gamma$  and in the next the generalized

eigenfunctions and the spectral representation of  $\Delta_\Gamma$ . The main new spectral characteristic which we must handle in the non-compact case is the absolutely continuous spectrum. We can see this already even when the manifold is flat.

*Example 2.10.* Let  $\mathcal{M} = \mathbb{R}^n$  with the Euclidean metric. The Laplacian  $\Delta_E = -\sum_{i=1}^n \frac{\partial^2}{\partial x_i^2} \geq 0$ . The spectral representation of  $\Delta_E$  is given via the Fourier transform on  $L^2(\mathbb{R}^n)$ :

$$(\mathcal{F}f)(k) \equiv \hat{f}(k) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} f(x) e^{-ix \cdot k} dx. \quad (2.8)$$

It is easy to check that  $\mathcal{F}$  diagonalizes the Laplacian, i.e.,

$$\mathcal{F} \Delta_E \mathcal{F}^{-1} = |k|^2. \quad (2.9)$$

This also shows that

$$\sigma(\Delta_E) = \sigma_{ac}(\Delta_E) = [0, \infty). \quad (2.10)$$

For comparison with latter work, let us write  $k \in \mathbb{R}^n$  as  $(\omega, \lambda) \in S^{n-1} \times \mathbb{R}_+$ , so that

$$\hat{f}(k) = \hat{f}(\omega, \lambda) \in L^2(S^{n-1} \times \mathbb{R}_+).$$

The functions appearing in (2.8) can be written as

$$e_0(x; \omega; \lambda) \equiv e^{-ix \cdot k} = e^{-i\lambda x \cdot \omega}, \quad (2.11)$$

and are eigenfunctions of  $\Delta_E$  in the sense that

$$\Delta_E e_0(x; \omega; \lambda) = \lambda^2 e_0(x; \omega; \lambda). \quad (2.12)$$

For each  $\omega \in S^{n-1} = \partial_\omega \mathbb{R}^n$ , there is one such eigenfunction with eigenvalue  $\lambda^2$ . These eigenfunctions are complete in the sense that the map

$$f \in L^2(\mathbb{R}^n) \rightarrow \hat{f}(\omega, \lambda) \equiv (2\pi)^{-n/2} \int_{\mathbb{R}^n} e_0(x; \omega; \lambda) f(x) dx \in L^2(S^{n-1} \times \mathbb{R}_+), \quad (2.13)$$

is an isometric isomorphism.

*Example 2.11.* Let  $\mathcal{M} = \mathbb{H}^n$  with the hyperbolic metric. The Laplacian is given as in (2.6) by

$$\Delta_{\mathbb{H}^n} \equiv -(x_n D_n)^2 + (n-1)x_n D_n + x_n^2 P, \quad (2.14)$$

where we write  $D_n \equiv \frac{\partial}{\partial x_n}$  and  $P = -\sum_{i=1}^{n-1} \frac{\partial^2}{\partial x_i^2}$  is the Euclidean Laplacian on  $\mathbb{R}^{n-1}$ . We

now construct the complete spectral theory for  $\Delta_{\mathbb{H}^n}$ . This will provide guidance for the third and fourth chapters when we construct the spectral theory for  $\Delta_\Gamma$  on  $\mathbb{H}^n/\Gamma$ . The construction of the complete spectral theory for  $\Delta_{\mathbb{H}^n}$  is a standard

1. *Separation of variables and solutions.* We want to construct the Green's function for  $\Delta_{\mathbb{H}^n}$ . To this end, we first solve

$$\Delta_{\mathbb{H}^n} u = \lambda u, \quad (2.15)$$

on  $L^2(\mathbb{H}^n, d\mu)$  where  $d\mu(x) = x_n^{-n} dx_1 \dots dx_n$ . We first change Hilbert spaces to  $L^2(\mathbb{R}^{n-1} \times \mathbb{R}_+, dx_1 \dots dx_n)$  by means of the unitary transformation  $U: L^2(\mathbb{R}^{n-1} \times \mathbb{R}_+, dV) \mapsto L^2(\mathbb{H}^n, d\mu)$  defined by

$$(Uf)(x) \equiv x_n^{n/2} f(x). \quad (2.16)$$

Transforming the Laplacian, we find

$$\tilde{\Delta}_{\mathbb{H}^n} \equiv U^{-1} \Delta_{\mathbb{H}^n} U = -x_n^2 D_n^2 - 2x_n D_n + x_n^2 P + \frac{n}{2} \left( \frac{n}{2} - 1 \right). \quad (2.17)$$

We now want to solve

$$\tilde{\Delta}_{\mathbb{H}^n} \tilde{u} = \lambda \tilde{u}. \quad (2.18)$$

Introducing the Fourier transform  $\hat{u}(\xi', x_n)$  with respect to  $x' \in \mathbb{R}^{n-1}$ ,

$$\tilde{u}(x', x_n) = (2\pi)^{-((n-1)/2)} \int_{\mathbb{R}^{n-1}} e^{ix' \cdot \xi'} \hat{u}(\xi', x_n) d^{n-1} \xi', \quad (2.19)$$

into the equation (2.18), we obtain

$$\left\{ -x_n^2 D_n^2 - 2x_n D_n + x_n^2 |\xi'|^2 + \frac{n}{2} \left( \frac{n}{2} - 1 \right) \right\} \hat{u}(\xi', x_n) = \lambda \hat{u}(\xi', x_n), \quad (2.20)$$

which is an ordinary differential equation. This is reduced to the modified Bessel's equation by the substitution

$$\hat{u}(\xi', x_n) = x_n^{-1/2} \hat{v}(\xi', x_n). \quad (2.21)$$

Writing  $\lambda = s(n-1-s)$  and  $s = \frac{n-1}{2} + i\sigma$ ,  $\sigma \in \mathbb{C}$ , we arrive at an equation for  $\hat{v}$ ,

$$\{x_n^2 D_n^2 + x_n D_n - [x_n^2 |\xi'|^2 + (i\sigma)^2]\} \hat{v} = 0. \quad (2.22)$$

This equation has two linearly independent solutions for all values of  $\sigma$ . These are modified Bessel functions,

$$K_{i\sigma}(|\xi'| x_n) \quad (2.23)$$

and

$$I_{i\sigma}(|\xi'| x_n). \quad (2.24)$$

The  $K_\nu$  function is regular at infinity;  $I_\nu$  is regular at zero. Furthermore,  $K_\nu$  has a singularity given by  $\{\Gamma(1-\nu)^{-1} (\frac{1}{2} |\xi'| x_n)^{-\nu} - \Gamma(1+\nu)^{-1} (\frac{1}{2} |\xi'| x_n)^\nu\}$  as the argument vanishes. To obtain a solution with a Fourier transform in  $\xi'$  for  $\text{Im } \sigma \geq 0$ , we take

$$|\xi'|^{i\sigma} K_{i\sigma}(|\xi'| x_n), \quad (2.25)$$



and find from (2.19) and (2.21) that

$$\tilde{u}(x', x_n) = (2\pi)^{-(n-1)/2} \int_{\mathbb{R}^{n-1}} e^{ix' \cdot \xi'} x_n^{-1/2} |\xi'|^{i\sigma} K_{i\sigma}(|\xi'| x_n) d^{n-1} \xi'. \quad (2.26)$$

For use below, we also note that for any  $w \in \mathbb{R}^{n-1}$ , the function

$$\tilde{u}(x', x_n; w) \equiv \tilde{u}(x' - w, x_n), \quad (2.27)$$

is also a solution of (2.18) with  $\lambda = s(n-1-s)$ . This is because  $e^{ix' \cdot w} \tilde{u}(\xi', x_n)$  also solves (2.18). Consequently, we obtain, for each  $w \in \mathbb{R}^{n-1}$ , a degenerate family of eigenfunctions with eigenvalue  $\lambda$  given by (2.26)–(2.27). Returning to  $L^2(\mathbb{H}^n, d\mu)$  via (2.16), these generalized eigenfunctions are

$$u(x', x_n; w) \equiv (2\pi)^{-(n-1)/2} x_n^{(n-1)/2} \int_{\mathbb{R}^{n-1}} e^{ix' \cdot (\xi' - w)} |\xi'|^{i\sigma} K_{i\sigma}(|\xi'| x_n) d^{n-1} \xi'. \quad (2.28)$$

2. *Green's function.* The kernel for the resolvent of  $\tilde{\Delta}_{\mathbb{H}^n}$  on  $L^2(\mathbb{R}^{n-1} \times \mathbb{R}_+, dV)$  can be constructed by means of the Fourier transform and the two independent solutions of (2.20) given by

$$x_n^{-1/2} \begin{cases} K_{i\sigma}(|\xi'| x_n) \\ I_{i\sigma}(|\xi'| x_n), \end{cases} \quad (2.29)$$

found above. By the standard construction, the Green's function for the ordinary differential operator on the left side of (2.20) is

$$\hat{G}((\xi', x_n), (\xi', y_n); s) = (x_n y_n)^{-1/2} \begin{cases} K_{i\sigma}(|\xi'| x_n) I_{i\sigma}(|\xi'| y_n); & x_n > y_n \\ K_{i\sigma}(|\xi'| y_n) I_{i\sigma}(|\xi'| x_n); & y_n > x_n \end{cases} \quad (2.30)$$

with  $s = \left(\frac{n-1}{2}\right) + i\sigma$ . This can be verified using the fact that the Wronskian is

$$W(K_v(z), I_v(z)) = K_v(z) I'_v(z) - K'_v(z) I_v(z) = 1/z. \quad (2.31)$$

Returning to  $L^2(\mathbb{H}^n, d\mu)$ , the kernel of the resolvent  $(\Delta_{\mathbb{H}^n} - s(n-1-s))^{-1}$  in  $L^2(\mathbb{H}^n, d\mu)$  is

$$G(x, y; s) = (2\pi)^{-(n-1)/2} (x_n y_n)^{n/2} \int_{\mathbb{R}^{n-1}} e^{i(x' - y') \cdot \xi'} \hat{G}((\xi', x_n), (\xi', y_n); s) d^{n-1} \xi', \quad (2.32)$$

with  $\hat{G}$  given in (2.30).

3. *Spectral density and spectra type.* One explicit formula (2.32) for the Green's function allows us write the spectral density for  $\Delta_{\mathbb{H}^n}$  by virtue of Stone's formula (cf. [RS1]). Recall that for a self-adjoint operator  $A$ , with resolvent  $R(z) = (A - z)^{-1}$ ,

and an interval  $[a, b]$ , this formula states that,

$$\frac{1}{2}(E_{[a,b]} + E_{(a,b)}) = \lim_{\varepsilon \rightarrow 0^+} \frac{1}{\pi} \int_a^b \operatorname{Im} R(E + i\varepsilon) dE, \quad (2.33)$$

in the strong sense. Letting  $E = \left(\frac{n-1}{2}\right) + \sigma^2$  and  $s = \frac{n-1}{2} + i\sigma$ , we find that,

$$\begin{aligned} G(x, y; s) - G(x, y; \bar{s}) &= \frac{2}{i\pi} (2\pi)^{-(n-1)/2} (x_n y_n)^{(n-1)/2} \sinh(\sigma\pi) \\ &\quad \times \int_{\mathbb{R}^{n-1}} e^{i(x' - y') \cdot \xi'} K_{i\sigma}(|\xi'| x_n) K_{i\sigma}(|\xi'| y_n), \end{aligned} \quad (2.34)$$

where we used the identity

$$K_v(z) = \frac{\pi}{2 \sin(v\pi)} [I_{-v}(z) - I_v(z)]. \quad (2.35)$$

Substituting into the right side of (2.33) yields

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0^+} \frac{1}{\pi} \int_a^b \operatorname{Im} R(E + i\varepsilon) dE &= (2\pi)^{-(n-1)/2} \pi^{-2} (x_n y_n)^{(n-1)/2} \int_{a'}^{b'} 2\sigma d\sigma \sinh(\sigma\pi) \\ &\quad \times \int_{\mathbb{R}^{n-1}} e^{i(x' - y') \cdot \xi'} K_{i\sigma}(|\xi'| x_n) K_{i\sigma}(|\xi'| y_n) d^{n-1} \xi', \end{aligned} \quad (2.36)$$

where  $\left(\frac{n-1}{2}\right) + (a')^2 = a$ , etc. Since the spectral family

$$\begin{aligned} F(\lambda) &\equiv (2\pi)^{-(n-1)/2} \pi^{-2} (x_n y_n)^{(n-1)/2} \int_0^\lambda 2\sigma d\sigma \sinh(\sigma\pi) \\ &\quad \times \int_{\mathbb{R}^{n-1}} e^{i(x' - y') \cdot \xi'} K_{i\sigma}(|\xi'| x_n) K_{i\sigma}(|\xi'| y_n) d^{n-1} \xi' \end{aligned} \quad (2.37)$$

is continuous in  $\lambda \in \mathbb{R}_+$ , we get that the kernel of the projection  $E_{(a,b)}$  for  $\Delta_{\mathbb{H}^n}$  is

$$\begin{aligned} E_{(a,b)}(x, y) &= (2\pi)^{-(n-1)/2} \pi^{-2} (x_n y_n)^{(n-1)/2} \int_{a'}^{b'} 2\sigma d\sigma \sinh(\pi\sigma) \\ &\quad \times \int_{\mathbb{R}^{n-1}} e^{i(x' - y') \cdot \xi'} K_{i\sigma}(|\xi'| x_n) K_{i\sigma}(|\xi'| y_n) d^{n-1} \xi'. \end{aligned} \quad (2.38)$$

This continuity shows that there are no embedded eigenvalues in the continuous spectrum. Furthermore, the absolute continuity of  $F$  implies that the spectral measure is absolutely continuous with respect to the Lebesgue measure. This indicates that the spectrum of  $\Delta_{\mathbb{H}^n}$  is purely absolutely continuous.

solution to the eigenvalue equation (2.15) given by

$$E_0(x; w; s) \equiv \pi^{-1} (2\pi)^{-(n-1)/2} x_n^{(n-1)/2} \int_{\mathbb{R}^{n-1}} e^{i(x' - w) \cdot \xi'} |\xi'|^{i\sigma} K_{i\sigma}(|\xi'| x_n) d^{n-1} \xi'. \quad (2.39)$$

We now show that these provide an eigenfunction expansion for  $\Delta_{\mathbb{H}^n}$ . We need two identities involving the modified Bessel functions, which are sometimes called Kontorovich–Lebedev Inversion Formulas. The first is

$$\frac{1}{\pi^2} \int_0^\infty 2\lambda \, d\lambda \sinh(\lambda\pi) K_{i\lambda}(u) K_{i\lambda}(v) = v\delta(u-v), \quad (2.40)$$

for  $u, v \in \mathbb{R}$ . The second is complementary,

$$[\sinh(\lambda u) \sinh(\lambda' u)]^{1/2} \frac{1}{\pi^2} \int_{\mathbb{R}} K_{i\lambda'}(u) K_{i\lambda}(u) \frac{du}{u} = \frac{\delta(\lambda - \lambda')}{2(\lambda\lambda')^{1/2}}. \quad (2.41)$$

We refer to [Le] and [T] for a proof of these identities. Paralleling the Fourier transform theory of Example 2.10, we introduce the map

$$\mathcal{F}: L^2(\mathbb{H}^n, d\mu) \mapsto \int_{(0, \infty)}^\oplus L^2(\mathbb{R}^{n-1}, d^{n-1}w)$$

by

$$(\mathcal{F}f)(\sigma; w) \equiv (2\pi)^{-(n-1)/2} \int_{\mathbb{H}^n} f(x) E_0(x; w; s) d\mu(x), \quad (2.42)$$

with  $s = \frac{n-1}{2} + i\sigma$ ,  $\sigma \in \mathbb{R}$ . Similarly, for  $g(\sigma; w) \in \int_{(0, \infty)}^\oplus L^2(\mathbb{R}^{n-1}, d^{n-1}w)$ , we define

$$(\mathcal{F}^{-1}g)(x) \equiv (2\pi)^{-(n-1)/2} \int_0^\infty \int_{\mathbb{R}^{n-1}} d^{n-1}w g(\sigma; w) \overline{E_0(x; w; s)} 2\sigma \sinh(\pi\sigma) d\sigma. \quad (2.43)$$

We want to prove that  $\mathcal{F}$  is an invertible isometry with inverse given by  $\mathcal{F}^{-1}$ . Since we have for  $f \in D(\Delta_{\mathbb{H}^n})$ ,

$$(\mathcal{F} \Delta_{\mathbb{H}^n} f)(\sigma, w) = s(n-1-s)(\mathcal{F}f)(\sigma, w) = \left[ \left( \frac{n-1}{2} \right)^2 + \sigma^2 \right] (\mathcal{F}f)(\sigma, w), \quad (2.44)$$

this verifies that  $\mathcal{H}' = \int_{(0, \infty)}^\oplus L^2(\mathbb{R}^{n-1}, d^{n-1}w)$  and  $\mathcal{F}$  provide a spectral representation for  $\Delta_{\mathbb{H}^n}$  and that

$$\sigma(\Delta_{\mathbb{H}^n}) = \left[ \left( \frac{n-1}{2} \right)^2, \infty \right) = \sigma_{ac}(\Delta_{\mathbb{H}^n}). \quad (2.45)$$

To prove that  $\mathcal{F}^{-1}\mathcal{F} = 1$ , we take  $g \in C_0^\infty(\mathbb{H}^n)$  and compute

$$\begin{aligned} (\mathcal{F}^{-1}\mathcal{F}g)(y) &= (2\pi)^{-(n-1)/2} \int_0^\infty 2\sigma \sinh(\sigma\pi) d\sigma \int_{\mathbb{R}^{n-1}} d^{n-1}w \overline{E_0(y; w; s)} (\mathcal{F}g)(\sigma, w) \\ &= (2\pi)^{-(n-1)} \int_0^\infty 2\sigma \sinh(\sigma\pi) d\sigma \int_{\mathbb{R}^{n-1}} d^{n-1}w \overline{E_0(y; w; s)} \int_{\mathbb{H}^n} g(x) E_0(x; w, s) \frac{d^n x}{x_n^n} \\ &= g(y). \end{aligned} \quad (2.46)$$

Here we used identity (2.40) and the standard representation for the delta function on  $\mathbb{R}^{n-1}$ . Since  $C_0^\infty(\mathbb{H}^n)$  is dense in  $L^2(\mathbb{H}^n, d\mu)$  this proves the first claim. Similarly, one uses (2.41) to show that  $\mathcal{F}^{-1}\mathcal{F} = 1$ . Hence,  $\mathcal{F}$  is an isomorphism. These calculations also show that  $\mathcal{F}$  is an isometry:

$$\|\mathcal{F}f\|_{\mathcal{H}'} = \|f\|_{L^2(\mathbb{H}^n, d\mu)}. \quad (2.47)$$

We note that we could have as well taken  $\int_{(-\infty, 0)}^\oplus L^2(\mathbb{R}^{n-1}, d^{n-1}w)$  as the spectral representation space and map  $\tilde{\mathcal{F}}$  from  $L^2(\mathbb{H}^n, d\mu)$  to this space. Hence, we obtain two unitary equivalent representations. We will show in chapter 4 how these eigenfunctions  $E_0(x; w; s)$  in (2.39) are related to the Eisenstein series. Note here, however, that  $\partial_\infty \mathbb{H}^n = \mathbb{R}^{n-1}$  plays the role of  $S^{n-1}$  in Example 2.10: it labels the degenerate eigenfunctions of  $\Delta_{\mathbb{H}^n}$  with common eigenvalue  $s(n-1-s)$ .

We can now state the main theorem of this section. It gives a general characterization of the spectrum of  $\Delta_\Gamma$  on  $\mathcal{M} = \mathbb{H}^n/\Gamma$ .

**Theorem 2.12.** *Let  $\Gamma$  be a discrete, torsion-free, geometrically finite subgroup of hyperbolic isometries of  $\mathbb{H}^n$ . Let  $\mathcal{M} = \mathbb{H}^n/\Gamma$  be the corresponding hyperbolic manifold with Laplacian  $\Delta_\Gamma$ . Assume that  $\mathcal{M}$  is non-compact. There are two cases.*

- (1)  *$\text{Vol } \mathcal{M} < \infty$ .  $\sigma_{\text{ess}}(\Delta_\Gamma) = \sigma_{\text{ac}}(\Delta_\Gamma) = [((n-1)^2, \infty)$ , with finitely-many eigenvalues of finite multiplicity in  $[0, ((n-1)/2)^2)$ . There are examples for which there are infinitely-many eigenvalues in  $[((n-1)/2)^2, \infty)$ .*
- (2)  *$\text{Vol } \mathcal{M} = \infty$ .  $\sigma_{\text{ess}}(\Delta_\Gamma) = \sigma_{\text{ac}}(\Delta_\Gamma) = [((n-1)/2)^2, \infty)$  with no embedded eigenvalues. There are at most finitely-many eigenvalues of finite multiplicity in  $[0, ((n-1)/2)^2)$ .*

Although this theorem is simple to state, its proof is rather involved. We will sketch the proof of the main parts of this theorem in the next few sections. We will only comment on the existence of infinitely-many embedded eigenvalues for certain  $\Gamma$  with  $\mathcal{M}$  of finite volume and on the absence of embedded eigenvalues in the infinite volume case. Another complete proof of Theorem 2.12 can be found in the papers of Lax and Phillips (which contain many other results). References to earlier work can be found there and in [FH], [FHP1–2].

Our proof of Theorem 2.12 is based on the method of weighted-estimates on the boundary-values of the resolvent as developed by Mourre. We begin with a review of this technique in the next section. We refer to [CFKS] for a textbook presentation with complete proofs.

The theory of local positive commutators as developed by E Mourre [Mo] is an important tool in the spectral analysis of elliptic operators. The abstract theory which we review here has been applied to Schrödinger operators, elliptic operators on manifolds [FH], and to the wave equation [DeBHS]. We will review the main points here, for the details we refer to [CFKS].

Let  $L$  be a self-adjoint operator on a Hilbert space  $\mathcal{H}$  whose spectrum we want to study. We associate abstract Sobolev spaces with  $L$  as follows. For  $s \geq 0$  define,

$$\mathcal{H}_s \equiv D((1 + |L|)^{s/2}), \quad \text{with the norm,}$$

$$\|\psi\|_s \equiv \|(1 + |L|)^{s/2}\psi\|_{\mathcal{H}}.$$

For  $s < 0$ , we define

$$\mathcal{H}_s \equiv \mathcal{H}_{-s}^*, \quad \text{the dual space.}$$

To analyse  $L$ , we suppose that there exists a skew-adjoint (i.e.  $A^* = -A$ ) operator  $A$  such that the pair  $(L, A)$  satisfy the following hypotheses,

(H1)  $D(A) \cap \mathcal{H}_2$  is dense in  $\mathcal{H}_2$ ;

(H2) the form  $[L, A] \equiv LA - AL$ , defined on  $D(A) \cap \mathcal{H}_2$  extends to a bounded operator mapping  $\mathcal{H}_2 \rightarrow \mathcal{H}_{-1}$ ;

(H3)  $\exists$  self-adjoint operator  $L_0$  with  $D(L_0) = D(L)$  such that the form  $[L_0, A]$  defined on  $D(L_0) \cap D(A)$  extends to a bounded operator from  $\mathcal{H}_{+2} \rightarrow \mathcal{H}$  and  $D(A) \cap D(L_0 A)$  is a core for  $L_0$ ;

(H4) the form  $[[L, A], A]$ , where  $[L, A]$  is defined in [H2], extends from  $\mathcal{H}_2 \cap D(LA)$  to a bounded operator from  $\mathcal{H}_2 \rightarrow \mathcal{H}_{-2}$ .

Hypothesis (H3) is technical and used to regularize certain operators. Note that the condition on  $[L_0, A]$  is weaker than that on  $[L, A]$  appearing in (H2). For the application to Schrödinger operators,  $L = -\Delta + V$  and  $L_0 = -\Delta$ . The operator  $A$  can often be taken to be  $A = \nabla \cdot x + x \cdot \nabla$ , the generator of dilations. For applications to scattering theory on hyperbolic manifolds with ends (see [FH] and [DeBHS]),  $L$  is the Laplace–Beltrami operator and  $L_0$  is a separable operator on each end. The operator  $A$  for a given  $L$  is called a *conjugate operator* for  $L$ .

#### DEFINITION 2.13

(the Mourre estimate) A self-adjoint operator  $L$  obeys a Mourre estimate (ME) on an interval on  $I \subset \mathbb{R}$  with conjugate operator  $A$  if  $A$  is skew-adjoint and (1)  $L$  and  $A$  satisfy (H1) and (H2), (2)  $\exists$  a constant  $\alpha$ ,  $0 < \alpha < \infty$ , and a compact operator  $K$ , such that if  $E_I(L)$  is the spectral projection for  $L$  and interval  $I$ , we have

$$E_I(L)[L, A]E_I(L) \geq \alpha E_I + K. \quad (2.48)$$

The first fruit of the ME is control of embedded eigenvalues.

**Theorem 2.14.** *If (H1)–(H3) hold for  $L$  and  $A$  and  $L$  satisfies a ME (2.48) on  $I$  with conjugate operator  $A$ , then  $L$  has finitely-many eigenvalues in  $I$  each with finite multiplicity.*

This result follows from the *ME* and the virial theorem. Without attention to technical questions concerning domains, suppose that  $E_0 \in I$  is an eigenvalue of infinite multiplicity with an orthonormal family  $\{\psi_i\}$  of eigenfunctions. Since  $E_I(L)\psi_i = \psi_i$ , we have a version of the virial theorem,

$$\begin{aligned} & \langle \psi_i, E_I[L, A]E_I\psi_i \rangle \\ &= \langle \psi_i, E_I[L - E_0, A]E_I\psi_i \rangle \\ &= \langle (L - E_0)\psi_i, AE_I\psi_i \rangle + \langle AE_I\psi_i, (L - E_0)\psi_i \rangle \\ &= 0. \end{aligned} \tag{2.49}$$

On the other hand, the *ME* (2.48) gives a lower bound on the matrix element

$$\langle \psi_i, E_I[L, A]E_I\psi_i \rangle \geq \alpha + \langle \psi_i, K\psi_i \rangle. \tag{2.50}$$

Since  $\psi_i \rightarrow 0$  weakly and  $K$  is compact,  $K\psi_i \rightarrow 0$  strongly and the right side of (2.50) is bounded below by  $\alpha/2$ . This contradicts (2.49). If  $L$  has infinitely many eigenvalues  $\{E_i\}$  in  $I$ , we can repeat the above argument using the corresponding eigenfunctions  $\{\psi_{i,k_i}\}_{i=1}^\infty$  where  $k_i$  labels the finite degeneracy of  $E_i$ . Let us note that if  $E \in I$  is an eigenvalue of  $L$ , the compact operator  $K$  serves to cancel the positive part  $\alpha E_I(L)$  since (2.48) will still hold at  $E$ .

Perhaps the most important application of the *ME* is to control the singular continuous spectrum of  $L$ . Let us recall the fundamental criteria for the absence of singular continuous spectrum for  $L$  in an interval  $I \subset \mathbb{R}$  (see [RS4]).

**Theorem 2.15.** *Suppose that for some  $p \geq 1$  and for all  $\phi \in \mathcal{D} \subset \mathcal{H}$ , with  $\mathcal{D}$  a dense set, we have*

$$\limsup_{\varepsilon \rightarrow 0} \int_I dE |\operatorname{Im} \langle \phi, (L - E - i\varepsilon)^{-1} \phi \rangle|^p < \infty, \tag{2.51}$$

*then  $\sigma_{\text{sc}}(L) \cap I = \emptyset$ .*

Hence, the main task is to control the boundary value of the resolvent. Mourre proved that the *ME* along with hypotheses (H1)–(H4) provide such control.

**Theorem 2.16.** *Suppose  $L$  and  $A$  satisfy (H1)–(H4). Then each point  $\lambda \notin \sigma_{\text{pp}}(L)$ , for which the *ME* holds in a small interval about  $\lambda$ , is contained in an open interval  $I$ , such that*

$$\limsup_{\delta \rightarrow 0^+} \left( \sup_{\mu \in I} \|(|A| + 1)^{-1} (L - \mu - i\delta)^{-1} (|A| + 1)^{-1} \| \right) < c, \tag{2.52}$$

*for a finite constant  $c > 0$ .*

**COROLLARY 2.17.**

*Under the hypothesis of Theorem 2.16, if the Mourre estimate holds on an interval  $I$ , then  $\sigma_{\text{sc}}(L) \cap I = \emptyset$ .*

*Proof.* For any  $\phi \in \mathcal{H}$ , set  $\psi_A \equiv (1 + |A|)^{-1} \phi$ . Then by Theorem 2.16,

$$\limsup_{\delta \rightarrow 0^+} |\langle \psi_A, (L - \mu - i\delta)^{-1} \psi_A \rangle| < c, \tag{2.53}$$

uniformly on  $I$ . Since  $\{\psi_A \equiv (1 + |A|)^{-1} \phi \mid \phi \in \mathcal{H}\} = D(A)$  is dense by (H1), the bound (2.53) implies that the fundamental criteria (2.51) holds on  $I$ .  $\square$

One can improve this theorem in two directions:

1. One can replace  $(1 + |A|)^{-1}$  by  $(1 + |A|)^{-\sigma}$  for any  $\sigma > \frac{1}{2}$ .
2. The boundary values  $(1 + |A|)^{-\sigma}(L - \mu - i0)^{-1}(1 + |A|)^{-\sigma}$ ,  $\sigma > \frac{1}{2}$ , exist as bounded operators and are Hölder continuous of order  $(\sigma - \frac{1}{2})(\sigma + \frac{1}{2})^{-1}$  in  $\mu$  on  $I$ .

The proof of Theorem 2.16 is quite deep and can be found in [CFKS]. The original proof of Mourre [Mo] uses a differential inequality (see [FS] for another proof).

Our goal is to apply Mourre theory to study the Laplace–Beltrami operator  $\Delta_\Gamma$  on  $\mathcal{M} = \mathbb{H}^n/\Gamma$ . To this end, we must construct a conjugate operator  $A$  for  $\Delta_\Gamma$ . This requires that we study the local geometry of  $\mathcal{M}$  at infinity, to which we now turn.

## 2.5 Geometry at infinity

In order to apply the Mourre theory as outlined in the last section, we must construct a conjugate operator  $A$  for  $\Delta_\Gamma$ . This construction depends on the analysis of certain model spaces which are isometric to neighborhoods of  $\mathcal{M}$  at infinity. We now describe in detail the asymptotic geometry of  $\mathcal{M}$ . As above,  $\Gamma$  is a discrete, geometrically finite subgroup of  $\text{Isom } \mathbb{H}^n$ , with no elliptic elements and such that  $\mathcal{M} = \mathbb{H}^n/\Gamma$  is non-compact.

**Theorem 2.18.** *The hyperbolic manifold  $\mathcal{M} = \mathbb{H}^n/\Gamma$  admits a decomposition of the form*

$$\mathcal{M} = K \cup \left( \bigcup_{i=1}^l U_i \right), \quad 1 \leq l < \infty, \quad (2.54)$$

where  $K$  is pre-compact in  $\mathcal{M}$ , and each  $U_i$  is a neighborhood of infinity isometric to an open subset of one of the following, each equipped with a hyperbolic metric,

- (1)  $B_+^n \equiv \{x \in \mathbb{H}^n \mid \sum_{i=1}^n x_i^2 < 1\}$ ; such a  $U_i$  is called a *regular neighborhood of infinity*;
- (2)  $V \times \mathbb{R}^+$ , where  $V$  is a flat,  $(n-1)$ -dimensional vector bundle over a torus (described below); such a  $U_i$  is called a *cuspidal neighborhood of infinity*.

Before sketching the proof of this proposition, let us give two motivating examples in  $\mathbb{H}^2$ .

**Example 2.19. Regular Neighborhood.** Let  $\Gamma = \langle \gamma_D \rangle$ , where  $\gamma_D$  is the dilation described in Example 1.20. We take  $\gamma_D z = \lambda z$ ,  $\lambda > 1$ . This is a hyperbolic subgroup; there are no parabolic fixed points. A fundamental domain is given by  $\{z \mid 1 < |z| < \lambda \text{ and } \text{Im } z > 0\}$ . There is a single closed geodesic at  $x_1 = 0$ ,  $1 \leq x_2 \leq \lambda$ . By the construction of  $\mathcal{M}$  outlined in § 1.7, it is easily seen that  $\mathcal{M}$  is topologically a cylinder:  $\mathbb{R} \times S^1$ . Let  $l = \log \lambda$ . Then a change of coordinates shows that the appropriate metric on  $\mathbb{R} \times S^1$  is given by

$$ds^2 = d\tau^2 + l^2 \cosh^2 \tau d\omega^2, \quad (2.55)$$

where  $\tau \in \mathbb{R}$  and  $\omega \in [0, 1)$ . The hyperbolic volume element  $dV$  becomes  $(l \cosh \tau) d\tau d\omega$ .

This shows that the cross-sectional area of  $S^1$  grows as  $|\tau| \rightarrow \infty$ . The manifold  $\mathcal{M}$  looks like a double-ended trumpet. We decompose  $\mathcal{M}$  as

$$\mathcal{M} = K_R \cup U_{1,R} \cup U_{2,R},$$

where, for any  $R > 0$ ,  $K_R$  is precompact and topologically the set  $(-R, R) \times S^1$ . The ends  $U_{i,R}$  are topologically  $[R, \infty) \times S^1$  with the metric (2.55). These are isometric with an open subset of  $\mathbb{H}^2$  given by

$$\{(x_1, x_2) \in \mathbb{H}^2 \mid -\frac{1}{2} < x_1 < \frac{1}{2}, \quad 0 < x_2 < C_R\}.$$

For all  $R$  sufficiently large, this set lies in  $B_+^2$ . Hence the two ends  $U_i$  are regular neighborhoods of infinity.

*Example 2.20. Cusp neighborhood.* The classic example of a cusp neighborhood is obtained with

$$\Gamma = \mathrm{PSL}(2, \mathbb{Z}) < \mathrm{PSL}(2, \mathbb{R}),$$

as described in Example 1.20. Although  $\gamma_I$  is elliptic (note that  $\gamma_I i = i$ ), it can be shown that  $\mathbb{H}^2/\Gamma$  is a manifold. We decompose  $\mathcal{M} = K \cup U_a$ , where  $U_a = [a, \infty) \times S^1$  with the metric

$$ds^2 = dr^2 + e^{-2r} dx^2, \quad (2.56)$$

obtained by setting  $r = \log x_2$ . Note that the volume form is  $e^{-r} dx dr$ , so that the cross-sectional area (radius of  $S^1$ ) is shrinking as  $r \rightarrow \infty$ . As we have seen,  $\mathrm{Vol}(\mathcal{M}) < \infty$ . The point  $\infty$  is a fixed point for the parabolic subgroup  $\langle \gamma_T \rangle$ , so  $\mathcal{M}$  has a cusp. The remaining set  $K \equiv \mathcal{M} \setminus ([a, \infty) \times S^1)$  is bounded and pre-compact. To see that  $U_a$  is a cusp neighborhood in the sense of Theorem 2.18, take  $V = S^1$ .

*Sketch of the Proof of Theorem 2.18.* 1. Because  $\Gamma$  is geometrically finite, we can find a fundamental domain bounded by finitely-many geodesic hypersurfaces. Hence, we can construct a finite open cover of  $F$  and extend it to  $\mathcal{M}$ . We take  $K$  to be a large, pre-compact region in  $F$ . The pre-compactness condition is satisfied as long as  $K$  is a positive Euclidean distance from  $\partial_\infty \mathbb{H}^n$ . We choose the ends  $U_i$  so that each  $U_i$  contains at most one parabolic fixed point in its Euclidean closure. If  $U_i$  contains no such fixed point, it is isometric to an open subset of  $B_+^n$  (by adjusting  $K$ , we can normalize  $U_i$  so it is isometric to an open subset of the unit semi-ball). If  $U_i$  contains a fixed point  $p_i$ , we need a separate analysis.

2. *Analysis of cusp neighborhoods.* This is somewhat involved and dimension dependent. We suppose that the neighborhood  $U_i$  contains (in its Euclidean closure) a fixed point  $p_i$  (if  $p_i = \infty$ , we consider the closure in the one-point compactification  $\hat{\mathbb{R}}^{n-1}$ ).

$n = 2$ . We can conformally map  $p_i$  to the point at infinity. Each cusp  $U_i$  is equivalent (by hyperbolic isometries) to  $[a_i, \infty) \times S^1$  with the metric (2.56), for some  $a_i > 0$ . The fact that there is one universal cusp in 2-dimensions is one of the simplifying features of the asymptotic geometry.

$n = 3$ . There are two types of cusps.



(i) *Maximal rank cusps (MR)*. In this case, the cusp  $p_i$  is isolated from the rest of  $\partial_\infty \mathcal{M}$ . Recall that  $\partial_\infty \mathcal{M}$  can be viewed as  $\partial_\infty \mathbb{H}^n / \Gamma$ , as  $\Gamma$  has a natural action on  $\mathbb{R}^{n-1}$ .  $U_i$  can be chosen so that  $\text{Vol } U_i < \infty$ . The model for  $U_i$  is  $[a, \infty) \times \mathcal{N}$ , where  $\mathcal{N}$  is a 2-dimensional compact manifold, that is, the two-torus  $\Pi^2$ . The metric is locally a warped product of the form

$$ds^2 = dr^2 + e^{-2r} g_{ij}^{\mathcal{N}}(x) dx_i dx_j, \quad (2.57)$$

where  $g_{\mathcal{N}}$  is a flat metric on  $\mathcal{N}$  and  $(x_1, x_2)$  are local coordinates on  $\mathcal{N}$ . The model is given by manifold  $\mathcal{M}_{\text{MR}} = \mathbb{H}^3 / \langle \gamma_1, \gamma_2 \rangle$ , where  $\{\gamma_1, \gamma_2\}$  generate the translations in  $\mathbb{R}^2$ ,

$$\gamma_1 x = (x_1 + 1, x_2, x_3), \quad (2.58)$$

$$\gamma_2 x = (x_1, x_2 + 1, x_3). \quad (2.59)$$

Clearly,  $\mathbb{R}^2 / \langle \gamma_1, \gamma_2 \rangle \cong \Pi^2$  and  $V = \Pi^2$  in the description of the theorem.

(ii) *Non-maximal rank cusp (NMR)*. For  $n = 3$ , there is one type of non-maximal rank cusp of rank 1. The model  $\mathcal{M}_{\text{NMR}} = \mathbb{H}^3 / \Gamma$ , where  $\Gamma$  has a rank one parabolic subgroup, i.e. a subgroup isomorphic with a translation group  $\langle \gamma_1 \rangle$  given in (2.58) in one-dimension. Topologically, the model space is  $S^1 \times \mathbb{R} \times \mathbb{R}^+$ . If  $\Gamma = \langle \gamma_1 \rangle$ , then  $\mathcal{M}_{\text{NMR}} = \mathbb{H}^3 / \langle \gamma_1 \rangle$  is this space with the hyperbolic metric. The vector bundle  $V$  of the theorem is trivial  $V = S^1 \times \mathbb{R}$ . NMR cusp neighborhoods in  $\mathbb{H}^3$  are isometric with an open subset of the slab  $S^1 \times \mathbb{R} \times \mathbb{R}^+$ . The fixed point is at  $\infty$ . Note that there are expanding  $(x_2)$  and contracting  $(x_1)$  directions as the fixed point is approached. The fixed point is not isolated from the remaining  $\partial_\infty \mathcal{M}$ . The boundary at infinity of  $\mathbb{H}^3 / \langle \gamma_1 \rangle$  consists of a cylinder pinching down at  $p$ . It is compact but has a singularity at  $p$ .

$n \geq 4$ . There are  $n - 1$  types of cusp neighborhoods.

(i) *NMR cusp (rank  $n - 1$ )*. This is the simplest cusp neighborhood. The neighborhood  $U_i$  is isometric to an open subset of  $[a_i, \infty) \times N_i$ , where  $N_i$  is a smooth  $(n - 1)$ -dimension, compact manifold. The simplest is  $\Pi^{n-1} = \mathbb{R}^{n-1} / \langle \gamma_1, \gamma_2, \gamma_3, \dots, \gamma_{n-1} \rangle$ , with each  $\gamma_i$  generating an independent translation (as in (2.58)–(2.59)). A general rank  $(n - 1)$ -Euclidean subgroup is more complicated. The classical Bieberbach theory gives a classification of these groups and their quotients. We refer the reader to [Chr] or the lecture notes of Thurston [Th].

(ii) *NMR cusp (rank  $< n - 1$ )*. We encounter non-trivial vector bundles  $V$  first in the case of  $n = 4$ . A cusp  $p$  has rank  $k$ ,  $1 \leq k \leq n - 2$ , if the parabolic subgroup fixing  $p$  contains  $k$  independent translations. These translations, however, may be coupled with discrete rotations. Consider the example in  $n = 4$  of a rank one cusp at  $p = \infty$  given by

$$\gamma(x_1, x_2, x_3, x_4) = (x_1 + 1, R(x_2, x_3), x_4), \quad (2.60)$$

where the map  $R: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is a rotation. This group  $\Gamma = \langle \gamma \rangle$  defines a manifold structure on  $[0, 1) \times \mathbb{R}^2 \times \mathbb{R}^+$  under the identification given by (2.60). The map  $(x_1, x_2, x_3) \rightarrow (x_1 + 1, R(x_2, x_3))$  on  $\mathbb{R}^3$  defines a vector bundle over  $S^1$  with fibre  $\mathbb{R}^2$ . The most interesting case occurs when the rotation  $R$  is through an irrational angle.

Then, the vector bundle cannot be finitely-covered by a trivial bundle. Analysis of these cusps is quite difficult and we refer the reader to [FHP3] for the details. In general the model space for a NMR cusp with rank  $k$  is

$$U_i \cong F \times \mathbb{R}^{n-k-1} \times \mathbb{R}^+, \quad (2.61)$$

where  $F$  is a compact manifold arising as the quotient of  $\mathbb{R}^k$  by a discrete Euclidean group action and  $\mathbb{R}^{n-k-1}$  constitutes the non-compact directions. As we have seen,  $F \times \mathbb{R}^{n-k-1}$  is the topological space of a non-trivial vector bundle  $V$ .  $\square$

The manifold that we have described can be compactified using a technique of Borel and Serre. These compactifications are described in the paper [MP]. These authors compute the spectrum of the Laplacian acting on forms and the cohomology groups for infinite volume, geometrically finite hyperbolic manifolds.

## 2.6 Model space estimates

We have introduced two types of model spaces in Proposition 2.18. The model space for a regular neighborhood of infinity is simply  $\mathbb{H}^n$  itself. The model space for a cusp neighborhood is  $V \times \mathbb{R}^+$ , where  $V$  is the flat vector bundle described in the proof of the proposition. Our program is to verify a Mourre estimate for each model space by constructing an appropriate model space conjugate operator. We will then glue these conjugate operators and model space  $ME$ 's together using a partition of unity and a localization formula. This philosophy was successfully used in the proof of a Mourre estimate for  $N$ -body Schrödinger operators, see [FHe1], [CFKS], [PSS].

Let  $\mathcal{M}_i$ ,  $i = 1, 2$ , represent the two types of model spaces.

$$\mathcal{M}_1 \equiv \mathbb{H}^n, \quad (\text{regular}) \quad (2.62)$$

$$\mathcal{M}_2 \equiv \mathbb{H}^n / \Gamma_k = V \times \mathbb{R}^+, \quad (\text{cusp}) \quad (2.63)$$

where  $\Gamma_k$  is a rank  $k$ ,  $1 \leq k \leq n-1$ , discrete subgroup of  $\text{Isom } \mathbb{H}^n$ . The cusp for the model spaces  $\mathcal{M}_2$  is at infinity. Recall that  $V = F_k \times \mathbb{R}^{n-k-1}$ , with some twisting induced by a discrete subgroup of rotations in  $n-k-1$  dimensions. The manifold  $F_k$  is compact. In local coordinates (which are global for  $\mathcal{M}_1$ ), the Laplacian is given by (2.6). We will write coordinates in the  $i$ -th model space  $\mathcal{M}_i$  as  $(x_{ij}) = (x_{i1}, \dots, x_{in})$ . We find it convenient to introduce the operator  $B_i \equiv (x_{in} D_n - c_n)$ , with  $c_n \equiv ((n-1)/2)$ . The Laplacian can be written as

$$\Delta_i \equiv -B_i^2 + x_n^2 P_i + c_n^2, \quad (2.64)$$

where  $P_i \equiv -\sum_{j=1}^{n-1} \frac{\partial^2}{\partial x_{ij}^2}$  in local coordinates.

There is a distinction between MR and NMR cusps which should be noted. The neighborhood  $U_i$  of a MR cusp has finite volume whereas the volume is infinite for a NMR cusp. The model space for a MR cusp is  $F \times \mathbb{R}^+$ , where  $F$  is a flat, compact,  $(n-1)$ -dimensional manifold. In the MR case, the operator  $P$  in (2.64) is simply the Laplacian  $\Delta_F$  on  $F$ , so  $\sigma(\Delta_F)$  is purely discrete and includes zero. In the NMR case,  $\sigma(P)$  is absolutely continuous and equals  $[0, \infty)$ . These differences are related to the fact that a finite volume, non-compact hyperbolic manifold may have embedded

separation of variables when  $V = F_k \times \mathbb{R}$  is a trivial bundle. The general case is proved in [FHP3].

The general spectral properties of  $\Delta_i$  are easy to establish since they can be treated by the method of separation of variables as in Example 2.11.

### PROPOSITION 2.21.

Let  $\mathcal{M}_i$  be a model space  $V_i \times \mathbb{R}^+$ , with  $V_1 = \mathbb{R}^{n-1}$  (regular) or  $V_2 = F_k \times \mathbb{R}^{n-k-1}$  (cusp of rank  $k$ ), with a hyperbolic metric. Then the Laplacian  $\Delta_i$  on  $L^2(\mathcal{M}_i, d\mu)$ , given in (2.64), has spectrum  $[((n-1)/2)^2, \infty)$  and is purely absolutely continuous, i.e.  $\sigma_{sc}(\Delta_i) = \phi$  and there are no eigenvalues.

*Proof.* For  $i=1$ , the absence of any eigenvalues and the absolute continuity of the spectrum was proved in Example 2.11. The cusp case is most easily treated by setting  $r = \log x_n$ . Since  $x_n > 0$ ,  $r \in \mathbb{R}$  and  $\mathcal{M}_2 = V_2 \times \mathbb{R}$ , with local coordinates  $(\omega, r)$ . Separating variables and changing Hilbert spaces to  $\tilde{\mathcal{H}} \equiv L^2(V_2 \times \mathbb{R}, d\omega dr)$ , where  $d\omega$  represents the local volume element on  $V_2$ , it is easy to show that  $\Delta_2$  is unitarily equivalent to

$$\tilde{\Delta}_2 = -D_r^2 + e^{2r}P + c_n^2,$$

acting on  $\tilde{\mathcal{H}}$ . The operator  $\tilde{\Delta}_2$  and the Hilbert space  $\tilde{\mathcal{H}}$  can be decomposed with respect to the spectral decomposition of  $P$ . In the MR case, this is a direct sum decomposition:

$$\tilde{\Delta}_2 \equiv \bigoplus_{k=1}^{\infty} (-D_r^2 + \lambda_k e^{2r} + c_n^2), \quad (2.65)$$

and

$$\tilde{\mathcal{H}} = \bigoplus_{k=1}^{\infty} L^2(\mathbb{R}, dr),$$

where  $\{\lambda_k\} = \sigma(P)$ . Using the methods of Example 2.11, one can show that  $\sigma(-D_r^2 + \lambda e^{2r} + c_n^2)$ , for any  $\lambda \in \mathbb{R}$ , is purely absolutely continuous (see also [Ti] pg. 93). This implies the result. The direct sum is replaced by a direct integral in the NMR cusp case and the analog of (2.65) is

$$\tilde{\Delta}_2 = \int_{(0, \infty)}^{\oplus} \Delta_\lambda d\lambda,$$

with

$$\Delta_\lambda \equiv -D_r^2 + \lambda e^{2r} + c_n^2,$$

acting on  $L^2(\mathbb{R}, dr)$ . The result again follows.  $\square$

We now turn to the construction of conjugate operators  $A_i$ ,  $i=1,2$ , for the two types of model spaces. Let  $\psi \in C^\infty(\mathbb{R})$  denote a monotone cut-off function such that

$$\psi(x) = \begin{cases} 0 & x > 1 \\ 1 & x < 1/2. \end{cases}$$

For  $S > 0$ , let  $\psi_S(x) \equiv \psi(x/S)$ . This parameter  $S$  will be useful in controlling derivatives. We introduce an unbounded operator on  $L^2(\mathcal{M}_i, d\mu)$  by

$$L_i \equiv \log(x_{\text{in}}^2 P_i). \quad (2.66)$$

Note that  $B_i$  and  $L_i$  are constructed from the vector fields  $x_{\text{in}} \frac{\partial}{\partial x_{ij}}$ ,  $j = 1, \dots, n$ . The conjugate operator  $A_i$  is defined by

$$A_i \equiv \psi_S(L_i)(L_i - 2S)B_i + B_i(L_i - 2S)\psi_S(L_i), \quad (2.67)$$

on an appropriate domain in  $L^2(\mathcal{M}_i, d\mu)$ .

Some motivation for this form of conjugate operator can be seen by studying a simple example. Let us consider the manifold  $\mathbb{R}^+ \times S^1$  with the Euclidean metric and the operator  $L_0 = -D_r^2 + e^{-r}P$ , with  $P = -\frac{d^2}{d\theta^2}$  on  $S^1$  and some self-adjoint boundary condition at  $r=0$  (which is not important here). As  $r \rightarrow \infty$ , this model mimics a regular neighborhood. Let us make some guesses for  $A$ . From Schrödinger operator theory, we first try  $A = \frac{1}{2}(rD_r + D_r r)$ , as in Remark 2.11. The commutator is

$$[L_0, A] = -2D_r^2 + re^{-r}P,$$

which is positive in the sense of the *ME* as  $r \rightarrow \infty$  but not relatively bounded. To make the commutator smaller we try  $A = D_r$ . This yields

$$[L_0, A] = e^{-r}P,$$

which is relatively bounded but not positive in the sense of the *ME*. Note that if  $P$  was a bounded operator, the first choice would be a good one for the remainder would be relatively compact. What we must do is balance the growth in  $\sigma(P)$  with the growth in  $r$ . To this end, we try

$$A = \frac{1}{2}[(r - \log P)D_r + D_r(r - \log P)].$$

(Note that if we write  $-r = \log x_n$  and consider  $x_n \rightarrow 0$ ,  $r - \log P = -\log(e^{-r}P) = -\log(x_n P)$  which appears in (2.66).) Now the commutator is

$$[L_0, A] = -2D_r^2 + (r - \log P)e^{-r}P. \quad (2.68)$$

The idea is to restrict to the region  $r - \log P > \delta > 0$  so that (2.68) is positive. Also, we have

$$(r - \log P)e^{-r}P = (r - \log P)e^{-(r - \log P)},$$

so it is bounded. This region is delimited by  $\psi_S$  given above. Finally, we must account for the region  $r - \log P < \delta$ . As we outline in the proof of Theorem 2.22, this region becomes disjoint from a neighborhood of a fixed energy as  $\delta$  becomes small.

We now verify the Mourre estimate for the model space Laplacians  $\Delta_i$  and the conjugate operator  $A_i$ .

**Theorem 2.22.** (Model space Mourre estimate). *For each  $\lambda > c_n^2 \equiv \left(\frac{n-1}{2}\right)^2$ , and for every  $\varepsilon > 0$ ,  $\exists$  a smoothed characteristic function  $f$  of some interval about  $\lambda$  such that for all  $S$  large enough*

$$f(\Delta_i)[\Delta_i, A_i]f(\Delta_i) \geq 8(\lambda - c_n^2 - \varepsilon)f(\Delta_i)^2.$$

*Sketch of the Proof*

1. *Technical estimates.* Let  $\mathcal{H}_s(\mathcal{M}_i)$  be the scale of spaces introduced in §2.4 relative to  $\Delta_i$ . One proves that

$$B_i: \mathcal{H}_s(\mathcal{M}_i) \rightarrow \mathcal{H}_{s-1}(\mathcal{M}_i), \quad s \in [-1, 2],$$

and that

$$x_{\text{in}} D_{ij}: \mathcal{H}_s(\mathcal{M}_i) \rightarrow \mathcal{H}_{s-1}(\mathcal{M}_i), \quad s \in [-1, 2],$$

and related maps are bounded. These are used to prove that

$$[\Delta_i, A_i]: \mathcal{H}_2(\mathcal{M}_i) \rightarrow L^2(\mathcal{M}_i),$$

is a bounded operator. Indeed, if we define

$$F_1(x) \equiv \psi(x) + \psi'(x)(x-2),$$

then from (2.67) one finds,

$$A_i = 2\psi_S(L_i)(L_i - 2S)B_i + 2F_1(L_i/S),$$

and a computation shows that

$$[\Delta_i, A_i] = -8B_iF_1(L_i/S)B_i + F_2(L_i/S),$$

where

$$F_2(x) \equiv -84\psi'''(x)(x-2)/S^2 - 24\psi''(x)/S^2 + 4\psi(x)S(2-x)e^{Sx}.$$

These are then shown to be bounded by the above estimates.

2. *The commutator.* We can compute a lower bound on the commutator from (2.69) as follows. Since  $\psi'(x)(x-2) \geq 0$ , we find that  $F_1(x) \geq \psi(x)$ , and

$$[\Delta_i, A_i] \geq -8B_i\psi_S(L_i)B_i + 4\psi_S(L_i)(-L_i - 2S)x_{\text{in}}^2P_i + F_4(L_i/S),$$

where  $F_4(x) = -8S^{-2}\psi'''(x)(x-2) - 24S^{-2}\psi''(x)$ . Now  $-x + 2S > S$  on  $\text{supp } \psi$  and  $\|F_4(L_i/S)\| = \mathcal{O}(S^{-2})$ , so for  $S$  large enough,

$$[\Delta_i, A_i] \geq 8(\Delta_i - c_n^2) - 8B_i(1-\psi)B_i - 8(1-\psi)x_{\text{in}}^2P_i + \mathcal{O}(S^{-2}) \quad (2.70)$$

3. *Positivity.* We choose a smooth function  $f \geq 0$  localized in  $[\lambda - \varepsilon, \lambda + \varepsilon]$ . Multiplying both sides of (2.70) by  $f(\Delta_i)$ , we obtain

$$f(\Delta_i)[\Delta_i, A_i]f(\Delta_i) \geq 8(\lambda - c_n^2 - \varepsilon)f(\Delta_i)^2 + f(\Delta_i)E(S)f(\Delta_i). \quad (2.71)$$

We must show that  $\|f(\Delta_i)E(S)\| \rightarrow 0$  as  $\varepsilon \rightarrow 0$  and  $S \rightarrow \infty$ . The intuition behind this is as follows. The main term in  $E(S)$  is  $1 - \psi_S(L_i)$ . This term localizes  $L_i$  to a spectral

region  $L_i > S/2$  or  $x_{in}^2 P > e^{S/2}$ . Since  $-B_i^2 \geq 0$ , this localizes the total energy  $(-B_i^2 + x_{in}^2 P_i + c_n^2) > e^{S/2}$ . On the other hand,  $f(\Delta_i)$  localizes the total energy near  $\lambda$ . It is clear that these two localization regions become disjoint as  $S \rightarrow \infty$  and  $\varepsilon \rightarrow 0$  and, consequently, the operator norm should approach zero.  $\square$

*Remark 2.23.* Note that there is no compact operator term in the model space Mourre estimate. Because of this, a virial type argument similar to the one used in the proof of Theorem 2.13, can be used to establish that the only eigenvalues of  $\Delta_i$  in  $[(n-1)/2)^2, \infty)$  can occur at  $((n-1)/2)^2$ . However, Theorem 2.22 gives no information about this point nor about the discrete spectrum of  $\Delta_i$  in  $[0, ((n-1)/2)^2)$ .

## 2.7 Global Mourre estimate

We now prove a ME for  $\Delta_\Gamma$ . Let  $\{U_0 \equiv K, U_1, \dots, U_l\}$  be the open cover of  $\mathcal{M}$  introduced in Proposition 2.18. Each  $U_i$  is isometric with an open subset of a model space  $\mathcal{M}_i$  (there are only two general types as discussed in the last section). We introduce a partition of unity  $\{\chi_i^2\}_{i=0}^l$  relative to that cover. We also need identification operators  $J_i$  so that

$$f \in C_0(U_i) \rightarrow J_i f \in C_0(\mathcal{M}_i). \quad (2.72)$$

One checks that  $J_i \chi_i: \mathcal{H}_s(\mathcal{M}) \rightarrow \mathcal{H}_s(\mathcal{M}_i)$  are bounded and related technical estimates. We will also use some compactness results. For example, the operators  $[\chi_i, f(\Delta_i)]$ ,  $f(\Delta)\eta$ , and  $(f(\Delta)J_i^* - J_i^* f(\Delta_i))\chi_i$  are shown to be compact for any  $f \in C^\infty(\mathbb{R}^n)$  vanishing at infinity and  $\eta$  defined after (2.77) below. By means of these operators, we define a conjugate operator  $A$  for  $\Delta_\Gamma$  on  $L^2(\mathcal{M}, d\mu)$ :

$$A = \sum_{i=1}^l \chi_i J_i^* A_i J_i \chi_i. \quad (2.73)$$

- This operator is localized on the neighborhoods of infinity of  $\mathcal{M}$ . The cut-off function  $\chi_0$  has compact support and  $\chi_0(\Delta_\Gamma + i)^{-1}$  is compact.

**Theorem 2.24. (Mourre Estimate).** *For each  $\lambda > c_n^2 = \left(\frac{n-1}{2}\right)^2$  and for every  $\varepsilon > 0$ ,  $\exists$  a smoothed characteristic function  $f$  of some interval about  $\lambda$  and a compact operator  $K$  such that for all  $S$  large enough,*

$$f(\Delta_\Gamma)[\Delta_\Gamma, A]f(\Delta_\Gamma) \geq 8(\lambda - c_n^2 - \varepsilon)f(\Delta_\Gamma)^2 + K. \quad (2.74)$$

*Sketch of the Proof.* We will use the identities

$$\Delta J_i^* \chi_i = J_i^* \Delta_i \chi_i, \quad (2.75)$$

and

$$\chi_i J_i \Delta = \chi_i \Delta_i J_i, \quad (2.76)$$

repeatedly. Moreover,  $\chi_i$  represents a localization operator on  $\mathcal{M}_i$  or  $\mathcal{M}$  as necessary. With this understanding,  $[\chi_i, J_i] = 0$ . We first examine

$$f(\Delta)J_i^*[\Delta_i, \chi_i A_i \chi_i]J_i f(\Delta) = f(\Delta)J_i^* \chi_i [\Delta_i, A_i] \chi_i J_i f(\Delta) + E_1 + E_2. \quad (2.77)$$

The remainder terms  $L_i$  are compact. This follows from the estimates mentioned above and the boundedness of  $\psi_S(L_i) L_i (\log x_{\text{in}})^{-1} \eta(x'_i, x_{\text{in}})$  on  $\mathcal{H}_S(\mathcal{M}_i)$  for  $\eta$  with compact support in  $x'_i$  and  $x_{\text{in}} < 1$ , the proof of which is technical. As for the first term in (2.77) we use the compactness result above to replace  $f(\Delta)$  by  $f(\Delta_i)$  and the model ME in Theorem 2.22 to obtain

$$\begin{aligned}
 & f(\Delta) J_i^* \chi_i [\Delta_i, A_i] \chi_i J_i f(\Delta) \\
 &= J_i^* f(\Delta_i) \chi_i [\Delta_i, A_i] \chi_i f(\Delta_i) J_i + K \\
 &= J_i^* \chi_i f(\Delta_i) [\Delta_i, A_i] f(\Delta_i) \chi_i J_i + K \\
 &\geq 8(\lambda - c_n^2 - \varepsilon) f(\Delta) \chi_i^2 f(\Delta) + K,
 \end{aligned} \tag{2.78}$$

where  $K$  is a compact operator (possibly changing from line to line). We have taken  $\varepsilon$  small and  $S$  large as required in Theorem 2.22. Finally, keeping in mind the identities (2.75)–(2.76), we have

$$\begin{aligned}
 f(\Delta) [\Delta, A] f(\Delta) &= \sum_{i=1}^l f(\Delta) J_i^* [\Delta_i, \chi_i A_i \chi_i] J_i f(\Delta) \\
 &\geq \sum_{i=1}^l 8(\lambda - c_n^2 - \varepsilon) f(\Delta) \chi_i^2 f(\Delta) + K \\
 &= 8(\lambda - c_n^2 - \varepsilon) (1 - \chi_0^2) f(\Delta)^2 + K \\
 &= 8(\lambda - c_n^2 - \varepsilon) f(\Delta)^2 + K,
 \end{aligned} \tag{2.79}$$

since  $\chi_0$  has compact support. □

## 2.8 The essential spectrum

It is well known that the essential spectrum of a self-adjoint operator on a non-compact manifold is determined by the behavior of the operator on functions supported in neighborhoods of infinity. Moreover, Weyl's theorem states that  $\sigma_{\text{ess}}$  is invariant under relatively compact perturbations. To compute  $\sigma_{\text{ess}}(\Delta_\Gamma)$ , we utilize the partition of unity subordinate to the cover  $\{K, U_1, \dots, U_l\}$  of  $\mathcal{M}$  constructed in the last section. Let  $\{j_i^2\}_{i=1}^l$  and  $j_0^2$  be such a  $C^1$ -partition of unity where  $j_0$  has compact support. The localized operator  $j_0 \Delta_\Gamma j_0$  is compactly supported, so by Weyl sequences it suffices to consider only the ends  $\{U_i\}_{i=1}^l$ .

We first determine the bottom of  $\sigma_{\text{ess}}(\Delta_\Gamma)$ . One can easily check the following form of the IMS localization formula (cf [CFKS]) for  $\Delta_\Gamma$ ,

$$\Delta_\Gamma = \sum_{i=0}^l (j_i \Delta_\Gamma j_i - |x_n \nabla j_i|^2) \equiv \tilde{\Delta}_\Gamma + K_0,$$

by computing  $[[\Delta_\Gamma, j_i], j_i]$  and summing. Because  $\text{supp } j_0 \subset K$ ,

$$K_0 \equiv j_0 \Delta_\Gamma j_0 - |x_n \nabla j_0|^2,$$

is the sum of a compactly supported and a relatively compact operator, so  $\sigma_{\text{ess}}(\Delta_\Gamma) = \sigma_{\text{ess}}(\tilde{\Delta}_\Gamma)$ . Furthermore, if there are no NMR cusps, we can choose the  $j_i$  so that

$\text{supp}|\nabla j_i|$  is compact. The situation is technically more difficult if there exist cusps of NMR since then  $\text{supp}|\nabla j_i|$  cannot be chosen to be compact. However, we can make a choice of  $j_i$ 's so that  $|\nabla j_i|$  decrease rapidly enough as  $x_n \rightarrow 0$  so that  $|\nabla j_i|(\Delta_\Gamma - z)^{-1}$ ,  $\text{Im } z \neq 0$ , is Hilbert-Schmidt. Consequently, it suffices to examine  $\inf \sigma_{\text{ess}} \left( \sum_{j=1}^l j_i \Delta_\Gamma j_i \right)$ .

Persson's formula (cf. [Agl]) is a convenient tool for this. It states that for a second-order self-adjoint elliptic operator  $L$  on a non-compact manifold  $\mathcal{M}$ ,

$$\inf \sigma_{\text{ess}}(L) = \sup_{\substack{N \subset \mathcal{M} \\ \text{compact}}} \left[ \inf_{\phi \in C_0^\infty(\mathcal{M} \setminus N)} \langle \phi, L\phi \rangle \|\phi\|^{-2} \right]. \quad (2.80)$$

Applying this to our sum, we see that for all  $\phi \in C_0^\infty(\mathcal{M} \setminus K)$ ,

$$\begin{aligned} \left\langle \phi, \sum_{i=1}^l j_i \Delta_\Gamma j_i \phi \right\rangle &= \sum_{i=1}^l \langle j_i \phi, \Delta_\Gamma j_i \phi \rangle \\ &\geq \inf_i \langle J_i j_i \phi, \Delta_i J_i j_i \phi \rangle, \end{aligned} \quad (2.81)$$

where  $J_i$  is defined in (2.72) and  $\Delta_i$  is the model space Laplacian. By Proposition 2.21, the right side of (2.81) is bounded below by  $\left( \frac{n-1}{2} \right)^2$ .

To prove that each  $E \in [((n-1)/2)^2, \infty)$  is in  $\sigma_{\text{ess}}(\Delta_\Gamma)$ , it suffices to construct a Weyl sequence in  $L^2(\mathcal{M})$  for  $E$  and  $\Delta_\Gamma$ . For any end  $U_i$ , we choose a sequence  $u_n \rightarrow 0$  weakly,  $\|u_n\| = 1$ , and for which  $\Delta_\Gamma u_i = \Delta_i J_i u_i \rightarrow E u_i$  strongly. The existence of such sequences for the model space operators can be proved given the analysis in the proof of Proposition 2.21. If  $\psi_0$  is an eigenfunction of  $P$  for eigenvalue zero, then  $\exp(-i\lambda r) g(r)\psi_0(x')$  is an approximate eigenfunction for the value  $E = c_n^2 + \lambda^2$ .

## 2.9 Outline of the Proof of Theorem 2.12

We can now outline the main aspects of the proof of Theorem 2.12. We established Corollary 2.7 that  $\Delta_\Gamma \geq 0$  and in §2.8 that  $\sigma_{\text{ess}}(\Delta_\Gamma) = [((n-1)/2)^2, \infty)$ . Furthermore, this implies that  $\sigma(\Delta_\Gamma) \cap [0, ((n-1)/2)^2]$  is discrete. We next apply Theorem 2.16 and Corollary 2.17 to conclude that  $\sigma_{\text{sc}}(\Delta_\Gamma) = \emptyset$ , that  $\sigma_{\text{ac}}(\Delta_\Gamma) = [((n-1)/2)^2, \infty)$ , and that there are at most isolated eigenvalues of finite multiplicity in  $[((n-1)/2)^2, \infty)$ . For this, we must verify (H1)–(H4). The estimates discussed in the previous sections suffice to establish (H1) and (H2) with a stronger conclusion:  $[\Delta_\Gamma, A]: \mathcal{H}_2(\mathcal{M}) \rightarrow L^2(\mathcal{M}, d\mu)$  is bounded. This allows us to take  $L = L_0 = \Delta_\Gamma$  in (H3). It remains to verify (H4). The bound on the double commutator is technical and somewhat laborious; we refer the reader to [FHP1]. When  $\text{Vol}(\mathcal{M}) < \infty$ , only maximal rank cusps can be present. In this case, the set of points in  $[((n-1)/2)^2, \infty)$  where the ME fails is discrete and countable. Consequently, the results on  $\sigma_{\text{sc}}(\Delta_\Gamma)$  and  $\sigma_{\text{ac}}(\Delta_\Gamma)$  are the same as in the infinite volume case.

It remains to consider the question of eigenvalues embedded in  $[((n-1)/2)^2, \infty)$ . When  $\text{Vol } \mathcal{M} = \infty$ , the absence of embedded eigenvalues has been proved by Lax and Phillips [LP1–4] and by Mazzeo [M] using Careleman-type estimates. It can also be proved using the conjugate operator presented here and the method of Froese–



Herbst [FHe2]. The case of finite volume is extremely subtle. Selberg [Se] proved that for  $\Gamma = \text{PSL}(2, \mathbb{Z})$ ,  $\mathbb{H}^2/\Gamma$  has infinitely-many embedded eigenvalues called cusp forms. The occurrence of such eigenvalues in the continuous spectrum seems to be quite rare. We will return to this in chapter 5. This concludes the sketch of the proof of Theorem 2.12.  $\square$

### 3. Spectral theory of non-compact hyperbolic manifolds

The goal of this section is to derive fine spectral properties of  $\Delta_\Gamma$ , where  $\Gamma < \text{Isom } \mathbb{H}^n$  is a discrete, geometrically finite subgroup of hyperbolic isometries without elliptic elements. Furthermore, we concentrate on the case when  $\mathbb{H}^n/\Gamma \equiv \mathcal{M}$  is non-compact. Although the results are valid in this general setting, the proofs when  $\Gamma$  contains parabolics are technically more difficult. For ease of exposition, we will assume that  $\Gamma$  contains no parabolics from §3.4 onward. Then, without loss of generality, we can work with  $n = 3$ .

By “fine spectral properties”, we mean the spectral representation of  $\Delta_\Gamma$  and an eigenfunction expansion. In order to obtain these, we (1) derive asymptotic expansions for the Green’s function  $G_\Gamma(u; x; s)$  on  $\mathcal{M}$  as one variable approaches  $\partial_\infty \mathcal{M}$ ; (2) study the limit

$$\lim_{x_n \rightarrow 0^+} x_n^{-s} G_\Gamma(u; x; s) \equiv E_\Gamma(u; x'; s),$$

which exists as a consequence of step 1, and show that  $E_\Gamma$  is a generalized eigenfunction of  $\Delta_\Gamma$ ; (3) derive a functional equation for the Green’s function  $G_\Gamma$  on the critical line  $\text{Re } s = (n-1)/2$ .

After these three steps, the spectral representation and eigenfunction expansion follow rather directly by general functional analysis arguments.

#### 3.1 Change of spectral parameter

Up until this point, we have been using  $z \in \mathbb{C}$  as the spectral parameter in the sense that we have written  $R(z) = (\Delta_\Gamma - z)^{-1}$ . Theorem 2.12 is expressed in this parametrization. For many reasons which will become apparent, it is useful to switch from  $z$  to a parameter  $s$  defined by the relation

$$z \equiv s(n-1-s), \tag{3.1}$$

with the principal branch of the square root. The appearance of  $(n-1)$  is related to the fact that  $\inf \sigma_{\text{ess}}(\Delta_\Gamma) = ((n-1)/2)^2$  in the  $z$ -parametrization. If we carry the results of Theorem 2.12 to the  $s$ -plane, we find that  $\sigma_{\text{ess}}(\Delta_\Gamma) = \{s \in \mathbb{C} \mid \text{Re } s = (n-1)/2\}$ . This is called the “critical line”. The effect of the transformation (3.1) is to open up the continuous spectrum  $[((n-1)/2)^2, \infty)$  onto the critical line. If  $z = ((n-1)/2)^2 \pm \sigma^2$ , then  $s = (n-1)/2 \pm i\sigma$ , for  $\sigma \in \mathbb{R}^+$ . This also shows that embedded eigenvalues become embedded in the critical line. The discrete spectrum in  $[0, ((n-1)/2)^2)$  is mapped to the interval  $((n-1)/2, n-1]$ . The resolvent of  $\Delta_\Gamma$  will now be written as

It is analytic on  $\operatorname{Re} s > ((n-1)/2)$ , except at those real  $s$  for which  $s(n-1-s) \in \sigma_d(\Delta_\Gamma)$ . The bottom of  $\sigma_{\text{ess}}(\Delta_\Gamma)$  corresponds to  $s = ((n-1)/2)$ .

### 3.2 Boundary at infinity

In §2.5, we described the geometry at infinity of  $\mathcal{M} \equiv \mathbb{H}^n/\Gamma$ . It will be important for later sections to have a description of  $\partial_\infty \mathcal{M}$ , the boundary of  $\mathcal{M}$  at infinity. The covering space  $\mathbb{H}^n$  has  $\partial_\infty \mathbb{H}^n = \hat{\mathbb{R}}^{n-1}$ , as described in §1.3. We saw there that  $\Gamma < \operatorname{Isom} \mathbb{H}^n$  induces an action on  $\hat{\mathbb{R}}^{n-1}$  and, in §1.6, we introduced the limit set  $\Lambda(\Gamma)$  and the domain of discontinuity  $\Omega(\Gamma) = \hat{\mathbb{R}}^{n-1} \setminus \Lambda(\Gamma)$  for this action. Then we have  $\partial_\infty \mathcal{M} = \Omega(\Gamma)/\Gamma$ , as a quotient manifold. It is possible to obtain a finite cover of  $\partial_\infty \mathcal{M}$  by extending the cover  $\{U_i\}_{i=1}^l$  of the ends of the manifold  $\mathcal{M}$ , given in Proposition 2.18, to  $\hat{\mathbb{R}}^{n-1}$ . We call these extensions  $U_i^\infty$  and note that  $\partial_\infty \mathcal{M} = \bigcup_{i=1}^l U_i^\infty$ . For calculational

purposes, we should think of these extensions in the upper half-space model, where they correspond  $\bar{U}_i \cap \partial_\infty \mathbb{H}^n$ . For a regular neighborhood,  $U_i^\infty$  is an open subset of a disk in  $\mathbb{R}^{n-1}$ . For a cusp neighborhood,  $U_i^\infty$  is an open subset of the bundle  $V$ . For  $n=3$ , this is a cylinder  $S^1 \times \mathbb{R}$ .

We describe  $\partial_\infty \mathcal{M}$  as follows, depending on the nature  $\Gamma$ ,

- (1)  $\Gamma$  purely hyperbolic.  $\partial_\infty \mathcal{M}$  is a smooth, compact  $(n-1)$ -dimensional manifold, possibly with several connected components. For  $\Gamma = \langle \gamma_0 \rangle$  in  $\mathbb{H}^3$ ,  $\partial_\infty \mathcal{M} = \Pi^2$ , a flat torus.
- (2)  $\Gamma$  has only MR parabolic subgroups.  $\partial_\infty \mathcal{M} = B \cup \{p_i\}_{i=1}^k$ , where  $B$  is a smooth, compact  $(n-1)$ -dimensional manifold as in (1) and the finite discrete set  $p_i \notin B$  are the maximal rank parabolic fixed points. As an example, consider the group  $\Gamma$  in  $\mathbb{H}^2$  generated by inversion (1.25) and translations (1.22) with  $b > 2$ . The manifold  $\mathbb{H}^2/\Gamma$  has one cusp at infinity and has infinity volume since  $b > 2$ . Then  $\partial_\infty \mathcal{M} = \{\infty\} \cup B$ , where  $B = S^1$ .
- (3)  $\Gamma$  arbitrary, geometrically finite discrete subgroup. As described in §2.5, the boundary  $\partial_\infty \mathcal{M}$  is complicated. It is, in general, the union of a compact manifold with singularities which are the NMR fixed points and finitely-many isolated MR fixed points. It is sometimes convenient to work with the non-compact component obtained by deleting the NMR fixed points. In  $\mathbb{H}^3$ , it follows from the proof of Proposition 2.18 that  $\partial_\infty \mathcal{M}$  is the union of disks (regular neighborhoods of infinity), half-cylinders and points (MR and NMR cusp neighborhoods, respectively).

One final remark concerning coordinate maps. As each  $U_i$  is isometric with an open subset of the model space  $\mathcal{M}_i$ , we introduce coordinate maps  $u_i: \mathcal{M}_i \rightarrow U_i \subset \mathcal{M}$ .

For regular neighborhoods, the domain of  $u_i$  is an open subset of  $\left\{w \in \mathbb{H}^n \mid \sum_{i=1}^n w_i^2 < 1\right\}$ ,

whereas for cusp neighborhoods, it is an open subset of  $V \times \mathbb{R}_+$ . The metric and Laplacian can be expressed locally in terms of these maps and the corresponding expressions are the same as above. We can extend these maps by taking  $x_n \rightarrow 0$  to maps on  $\mathbb{R}^{n-1}$  with range  $U_i^\infty$ . We call the boundary maps  $u_i^\infty$ . When  $\mathcal{M}$  is fixed, we will often write  $B$  for  $\partial_\infty \mathcal{M}$ .

### 3.3 Localization formulas for the resolvent

In order to study the asymptotic properties of the Green's function, we localize the resolvent in the neighborhoods of infinity  $U_i$ . Let  $u_i$  be the coordinate maps introduced above. Let  $\mathcal{B}(X)$  denote the bounded functions on  $X$ . We introduce identification maps  $J_i: \mathcal{B}(U_i) \rightarrow \mathcal{B}(\mathcal{M}_i)$  by

$$(J_i f)(w) \equiv f(u_i(w)), \quad (3.3)$$

where we write  $w \in \mathcal{M}_i$  as above. The adjoint map  $J_i^*: \mathcal{B}(\mathcal{M}_i) \rightarrow \mathcal{B}(U_i)$ . Let  $\chi_i$  be a smoothed characteristic function on  $U_i$ . Then  $J_i \chi_i$  is such a function on  $\mathcal{M}_i$ . To simplify the notation, we write  $\chi_i$  for  $J_i \chi_i$ . Note that we also have

$$[\Delta_\Gamma, \chi_i] = [\Delta_i, \chi_i],$$

as the reader can verify. We now give formulas expressing the localization of  $(\Delta_\Gamma - z)^{-1}$  to the neighborhoods  $U_i$  in terms of  $(\Delta_i - z)^{-1}$ . These formulas are independent of the present geometric setting and have many other applications.

#### Lemma 3.1

1. Suppose that  $\chi_1$  and  $\chi_2$  have supports in  $\mathcal{M}$ . Then on the model space  $\mathcal{M}_2$ ,

$$\begin{aligned} J_1 \chi_1 (\Delta_\Gamma - z)^{-1} \chi_2 J_2^* \\ = -(\Delta_1 - z)^{-1} [\Delta_1, \chi_1] J_1 (\Delta_\Gamma - z)^{-1} J_2^* [\Delta_2, \chi_2] (\Delta_2 - z)^{-1}, \end{aligned} \quad (3.4)$$

for  $z \in \rho(\Delta_\Gamma) \cap \rho(\Delta_i)$ ,  $i = 1, 2$ .

2. If  $\chi_1 = \chi_2$ , then on the model space  $\mathcal{M}_i$ ,

$$\begin{aligned} J_1 \chi_1 (\Delta_\Gamma - z)^{-1} \chi_1 J_1^* \\ = \chi_1 (\Delta_1 - z)^{-1} \chi_1 + (\Delta_1 - z)^{-1} [\Delta_1, \chi_1] (\Delta_1 - z)^{-1} [\Delta_1, \chi_1] (\Delta_1 - z)^{-1} \\ - (\Delta_1 - z)^{-1} [\Delta_1, \chi_1] J_1 (\Delta - z)^{-1} J_1^* [\Delta_1, \chi_1] (\Delta_1 - z)^{-1}. \end{aligned} \quad (3.5)$$

*Proof.* We begin with a formula valid on functions on  $\mathcal{M}_i$ ,

$$\chi_i J_i^* (\Delta_i - z) = \chi_i (\Delta_\Gamma - z) J_i^*, \quad (3.6)$$

which follows from the definition of local coordinates. Multiply (3.6) on the left by  $(\Delta_\Gamma - z)^{-1}$  and on the right by  $(\Delta_i - z)^{-1}$ , to obtain

$$\begin{aligned} (\Delta_\Gamma - z)^{-1} \chi_i J_i^* &= (\Delta_\Gamma - z)^{-1} \chi_i (\Delta_\Gamma - z) J_i^* (\Delta_i - z)^{-1} \\ &= \chi_i J_i^* (\Delta_i - z)^{-1} + (\Delta_\Gamma - z)^{-1} [\chi_i, \Delta_i] J_i^* (\Delta_i - z)^{-1}. \end{aligned} \quad (3.7)$$

Take the adjoint of (3.7) and set  $i = 1$ . Multiplying the resulting equation by  $\chi_i J_i^*$  on the right, we get

$$\begin{aligned} J_1 \chi_1 (\Delta_\Gamma - \bar{z})^{-1} \chi_i J_i &= (\Delta_1 - \bar{z})^{-1} J_1 \chi_1 \chi_i J_i^* \\ &\quad + (\Delta_1 - \bar{z})^{-1} J_1 [\Delta_1, \chi_1] (\Delta_\Gamma - \bar{z})^{-1} \chi_i J_i^*. \end{aligned} \quad (3.8)$$

(3.8). The results (3.4)–(3.5) follow from this.  $\square$

The basic idea of the next few sections is this. Formulas (3.4) and (3.5) allow us to express the behavior of  $(\Delta_\Gamma - z)^{-1}$  in the neighborhoods of infinity in terms of the resolvents of the model space Laplacians  $\Delta_i$ , provided we can control  $(\Delta_\Gamma - z)^{-1}$  as an operator. This was achieved in §2.7 where we derived *a priori* bounds on the weighted boundary values of  $(\Delta_\Gamma - z)^{-1}$ . We are then left with the task of studying the model space Green's functions. This is possible since they are explicitly computable.

### 3.4 Localization formulas for the Green's function

We now translate the formulas of Lemma 3.1 into expressions for the Green's function  $G_\Gamma(x, w; s)$ . From this point onward, we will assume that there are no cusps. We refer to [FHP2, 3] for the general case. Let  $A_M$  be the conjugate operator constructed in §2.7. We proved as a consequence of Theorem 2.12 and Theorem 2.16 that the operator

$$R_M(s) \equiv (|A_M| + 1)^{-1} (\Delta_\Gamma - s(n - 1 - s))^{-1} (|A_M| + 1)^{-1}, \quad (3.9)$$

is analytic for  $\operatorname{Re} s > (n - 1)/2$ , except at poles where  $s(n - 1 - s) \in \sigma_d(\Delta_\Gamma)$ , and has continuous boundary values on  $\operatorname{Re} s = (n - 1)/2$  ( $s \neq (n - 1)/2$ ). Let  $\{\chi_i\}$  be the partition of unity introduced in §3.3 with compatible coordinate functions  $u_i: \mathcal{M}_i \rightarrow U_i$ , as in §3.2. Let  $J_i: \mathcal{B}(U_i) \rightarrow \mathcal{B}(\mathcal{M}_i)$  be as defined in (3.3). Finally, let  $G_i$  be the Green's function for the model space Laplacian  $\Delta_i$  in the  $s$ -parametrization (cf. (3.2)).

*Lemma 3.2. Suppose that  $\chi_1$  and  $\chi_2$  have disjoint supports and  $x \in \{y | \chi_1(y) = 1\}$  and  $w \in \{y | \chi_2(y) = 1\}$ . Then,*

$$G_\Gamma(u_1(x); u_2(w); s) = \langle H_{x,s}^{(1)}, J_1 R_M(s) J_2^* H_{w,s}^{(2)} \rangle, \quad (3.10)$$

where, if  $\chi_i$  has support in a neighborhood of infinity,

$$H_{w,s}^{(i)}(\cdot) = (\tilde{\chi}_i A_i + 1) [\Delta_i, \chi_i] G_i(\cdot; w; s), \quad (3.11)$$

and, if  $\chi_i$  has compact support,

$$H_{w,s}^{(i)}(\cdot) = [\Delta_i, \chi_i] G_i(\cdot; w; s).$$

We also have

$$\frac{\partial G_\Gamma}{\partial w_3}(u_1(x); u_2(w); s) = \left\langle H_{x,s}^{(1)}, J_1 R_M(s) J_2^* \frac{\partial H_{w,s}^{(2)}}{\partial w_3} \right\rangle. \quad (3.12)$$

If  $x, w \in \{y | \chi_1(y) = 1\}$ , then we have

$$G_\Gamma(u_1(x); u_2(w); s) = G_1(x; w; s) + \langle H_{x,s}^{(1)}, (-R_{M_1}(s) + J_1 R_M(s) J_2^*) H_{w,s}^{(1)} \rangle.$$

Here,  $\langle \cdot, \cdot \rangle$  denotes the inner products on  $L^2(\mathcal{M}_i)$ .

This lemma follows directly from Lemma 3.1 provided the vectors  $H_{w,s}^{(i)}$  and  $\partial H_{w,s}^{(i)} / \partial w_3$  are in  $L^2(\mathcal{M}_i)$ , for then  $J_i^* H_{w,s}^{(i)} \in L^2(\mathcal{M})$ , etc. Furthermore, as we are

interested in the asymptotic behavior of  $G_\Gamma$ , formula (3.10) indicates that this will follow from an  $L^2$ -asymptotic expansion of  $H_{w,s}^{(i)}$  as  $w_3 \rightarrow 0$ . Since the Green's functions  $G_i$  for the model space is explicitly computable, we will be able to obtain these expansions of  $H_{w,s}^{(i)}$  and  $\partial H_{w,s}^{(i)}/\partial w_3$ .

### 5 Model space Green's function estimates

We now turn to the classical analysis of special functions to obtain  $L^2$  asymptotic expansions of  $H_{w,s}^{(i)}$  and its derivatives. Because of the definition (3.11), we need expansions of  $G_i$  and its covariant derivatives. The following lemma holds for cusp neighborhoods also. For  $w = (w_1, w_2, w_3)$ , we will denote  $(w_1, w_2) \in \mathbb{R}^2$  by  $w'$ .

**Lemma 3.3.** *Let  $G_i(z; w; s)$  be a model space Green's function. Let  $z$  and  $w$  belong to disjoint sets, each of which is bounded in the Euclidean metric and separated from each other in the Euclidean metric by a positive distance. Then there exist bounded functions  $g_{w',s}^m(z)$  and, for  $m = 1, 2, 3$ ,  $g_{w',s}^m(z)$  and  $h_{w',s}^m(z)$ , analytic in  $s$  for  $\operatorname{Re} s > 1$ , continuous in  $s$  for  $\operatorname{Re} s \geq 1$  and  $\operatorname{Re} s \neq 1$  and smooth in  $w'$  so that for  $w_3 \rightarrow 0$ , we have,*

$$G_i(z; w; s) = z_3^s w_3^s (g_{w',s}^0(z) + \mathcal{O}(w_3)), \quad (3.13)$$

$$\frac{\partial G_i}{\partial w_3}(z; w; s) = s z_3^s w_3^{s-1} (g_{w',s}^1(z) + \mathcal{O}(w_3)),$$

$$z_3 \frac{\partial G_i}{\partial z_m}(z; w; s) = \begin{cases} w_3^s z_3^{s+1} (g_{w',s}^m(z) + \mathcal{O}(w_3)), & m = 1, 2 \\ s w_3^s z_3^s (g_{w',s}^3(z) + \mathcal{O}(w_3)), & m = 3 \end{cases}$$

$$\frac{\partial}{\partial w_3} z_3 \frac{\partial G_i}{\partial z_m}(z; w; s) = \begin{cases} s w_3^{s-1} z_3^{s+1} (h_{w',s}^m(z) + \mathcal{O}(w_3)), & m = 1, 2 \\ s^2 w_3^{s-1} z_3^s (h_{w',s}^3(z) + \mathcal{O}(w_3)), & m = 3. \end{cases} \quad (3.14)$$

the remainders are uniform in  $z, w'$ , and  $s$ .

**Sketch of the Proof.** We treat the case when the model space is  $\mathbb{H}^3$ . For any  $n$ , the Green's function for  $\mathbb{H}^n$  can be expressed as

$$G(z, w; s) = c(s) \sigma(z, w)^{-s} F_s(\sigma^{-1}(z, w)), \quad (3.15)$$

where

$$c(s) = (2s - n - 1)^{-1} \frac{2^{n-1-2s} \Gamma(s)}{\pi^{(n-1)/2} \Gamma(s - (n-1)/2)},$$

$$F_s(v) = {}_2F_1(s; s - (n/2) + 1; 2s - n + 2; v),$$

and

$$\sigma^{-1}(z, w) = 4z_n w_n (\|z' - w'\|_E^2 + (w_n + z_n)^2)^{-1}.$$

Here,  ${}_2F_1$  is the hypergeometric function. The function  $\sigma(z, w)$  is related to the hyperbolic distance from  $z$  to  $w$  as in (1.3). We now take  $n = 3$  although a similar argument holds in  $n$  dimensions. Let  $h(z, w) \equiv \|z' - w'\|_E^2 + (z_3 + w_3)^2$ . By assumption,

$$h(z, w) > \delta,$$

so  $h^{-1}$  exists and is  $C^\infty$  in the region specified. Moreover its derivatives satisfy

for some constants  $c_{\alpha\beta\delta}$ . The first order Taylor expansion of  $G$  about  $\sigma^{-1} = 0$  gives

$$G(z, w; s) = c(s)(w_3 z_3)^s h^{-s}(z, w) + (z_3 w_3)^{s+1} h^{-(s+1)}(z, w) \psi\left(\frac{4z_3 w_3}{h(z, w)}\right), \quad (3.16)$$

where  $\psi_s(x)$  is analytic in  $x$  in the given region and in  $s$  as determined by  $F_s(x)$ . Hence we see that

$$G(z, w; s) = z_3^s w_3^s g(z, w; s), \quad (3.17)$$

with  $g$  analytic in  $(z, w)$  down to  $z_3 = w_3 = 0$  as long as  $h > \delta$ . The statements of the lemma now follow from (3.16) and (3.17).  $\square$

We now prove that  $H_{w,s}^{(i)}$  and  $\partial H_{w,s}^{(i)}/\partial w_3$  are  $L^2$ -functions and obtain their  $L^2$ -asymptotic expansions. Since we are restricting ourselves to a regular neighborhood, we drop the superscript  $i$ .

*Lemma 3.4.* *Let  $\chi$  be a smoothed characteristic function of a regular neighborhood as in §3.3. For  $w'$  in the region where  $\chi = 1$ , but disjoint from the region  $\text{supp } \nabla \chi$ , there exist an  $L^2$ -function  $h_{w',s}$ , smooth in  $w'$  and analytic in  $s$  for  $\text{Re } s > 1$  and continuous for  $\text{Re } s \geq 1$ ,  $s \neq 1$ , s.t.*

$$\|H_{w,s} - w_3^s h_{w',s}\| \leq c w_3^{\text{Re } s + 1} \quad (3.18)$$

and

$$\left\| \frac{\partial H_{w,s}}{\partial w_3} - s w_3^{s-1} h_{w',s} \right\| \leq c w_3^{\text{Re } s}, \quad (3.19)$$

on the model space  $L^2(\mathcal{M}_i)$ .

*Sketch of the Proof.* Recall the definition of  $H_{w,s}$  given in (3.11) and of  $A_i$  given in (2.67). Then we must consider

$$[\Delta_\Gamma, \chi] = -2z_3 \frac{\partial \chi}{\partial z_3} B - z_3 \frac{\partial}{\partial z_3} \left( z_3 \frac{\partial^2 \chi}{\partial z_3} \right) - 2 \sum_{m=1}^2 \left( z_3 \frac{\partial \chi}{\partial z_m} (z_3 D_m) + z_3^2 \frac{\partial^2 \chi}{\partial z_m^2} \right),$$

where  $B = x_3 D_3$ , and

$$\tilde{\chi} A_i + 1 = \tilde{\chi} (2\xi(L/S)(L-2S)B + 2F_1(L/S)) + 1,$$

where  $\xi$ ,  $L$  and  $F_1$  are defined in §2.6, respectively. In §2.7, we proved the  $L^2$ -boundedness of the following operators,

$$\xi(L/S) L(\log z_3)^{-1} \eta, \text{ with } \eta = 1 \text{ on } \text{supp } \nabla \chi,$$

$$B^2(\Delta + 1)^{-1},$$

$$z_3 D_m(\Delta + 1)^{-1},$$

so we can write

$$H_{w,s}(z) = C\omega\eta(\Delta + 1)G(z; w; s) + C[\Delta, \omega\eta]G(z; w; s)$$

where

$$\omega \equiv \omega(z_3) = z_3 \log z_3$$

and  $C$  is a bounded operator. Hence, we need to obtain  $L^2$ -expansions for

$$\omega(z_3)\eta(z)G(z, w; s)$$

and

$$\omega(z_3)\eta(z)z_3 \frac{\partial G}{\partial z_m}(z, w; s) \quad m = 1, 2, 3.$$

These follow from the expansions of  $G$  given in Lemma 3.3. Of particular note is the fact that for  $\operatorname{Re} s \geq 1$ ,

$$\int_0^1 \omega(z_3)^2 (z_3^{2\operatorname{Re} s}) (z_3^{-3}) dz_3 < \infty,$$

which is needed to insure that the resulting norms are finite. The proof for  $\partial H_{w,s}/\partial w_3$  is similar.  $\square$

### 3.6 $L^2$ -Asymptotic expansions of the Green's function

We can now use the results of Lemma 3.4 and the formulas in Lemma 3.2 to obtain the  $L^2$ -asymptotic expansions of the Green's function  $G_\Gamma(u; v; s)$  of  $\Delta_\Gamma$ . This is the first main result of this chapter.

**Theorem 3.5.** *Let  $\Gamma < \operatorname{Isom} \mathbb{H}^3$  be a discrete, geometrically finite subgroup of hyperbolic isometries without elliptic or parabolic elements. Let  $G_\Gamma(u; v; s)$  be the Green's function of  $\Delta_\Gamma$  defined initially for  $\operatorname{Re} s > 1$ ,  $u \neq v$ , and  $s(2-s) \notin \sigma_d(\Delta_\Gamma)$ . Let  $(w_1, w_2, w_3)$  be a coordinate system for a regular neighborhood of infinity with coordinate functions  $u_i$ . Then for  $\operatorname{Re} s \geq 1$ ,  $s \neq 1$ , and  $s(2-s) \notin \sigma_d(\Delta_\Gamma)$ ,  $G_\Gamma$  has the following asymptotics,*

$$G_\Gamma(u; u_i(w); s) = w_3^s E_i(u; w'; s) + \mathcal{O}(w_3^{\operatorname{Re} s + 1}) \quad (3.20)$$

$$\frac{\partial G_\Gamma}{\partial w_3}(u; u_i(w); s) = s w_3^{s-1} E_i(u; w'; s) + \mathcal{O}(w_3^{\operatorname{Re} s}). \quad (3.21)$$

The error terms are uniform in  $u, w'$  and  $s$  in compact sets. Then functions  $E_i(u; w'; s)$  are smooth in  $u$  and  $w'$  and analytic in  $s$  for  $\operatorname{Re} s > 1$ ,  $s(2-s) \notin \sigma_d(\Delta_\Gamma)$  and continuous onto  $\operatorname{Re} s = 1$ , except possibly at  $s = 1$ .

*Proof.* Let us consider  $u$  in a compact subset of  $\mathcal{M}$  whereas  $w$  approaches  $\partial_\infty \mathbb{H}^3$  with  $w'$  remaining in a compact set. Then  $u_i(w)$  remains in  $U_i$  and approaches  $\partial_\infty \mathcal{M}$ . We choose localization functions  $\chi_1$  with compact support in  $\mathcal{M}$  and  $\chi_i$  in  $U_i$ . We can then use formulas (3.13)–(3.14) for  $G_\Gamma$  and  $\partial G_\Gamma/\partial w_3$ . The vectors  $H_{x(u),s}^{(i)}$  are given in Lemma 3.4. The result for  $G_\Gamma$  is now immediate with the function  $E_i$  defined by

$$E_i(u; w', s) \equiv \langle H_{x(u),s}^{(1)}, J_1 R_{\mathcal{M}}(s) J_i^* h_{w',s}^{(i)} \rangle, \quad (3.22)$$

where  $x_j(u)$  are the local coordinate functions for  $u \in \operatorname{supp} \chi_1$ . The expression for  $\partial G_\Gamma/\partial w_3$  is obtained in the same way.  $\square$

### 3.7 Generalized eigenfunctions for $\Delta_\Gamma$

We now study the functions  $E_i(u; w'; s)$  appearing in Theorem 3.5. For fixed  $u \in \mathcal{M}$ , these functions are locally defined on  $U_i^\infty$ . As functions of  $u \in \mathcal{M}$ , they are, in fact, eigenfunctions of  $\Delta_\Gamma$  although they are not in  $L^2(\mathcal{M})$ . We now expand upon these two points.

(a)  $E_i(u; w'; s)$  is an eigenfunction of  $\Delta_\Gamma$  with eigenvalue  $s(2-s)$ ,  $\operatorname{Re} s = 1$ ,  $s \neq 1$ . From the definition of  $G_\Gamma$  and the existence of its continuation onto the critical line, we have in the variable  $u \in \mathcal{M}$ ,

$$(\Delta_\Gamma - s(2-s))G_\Gamma(u; u_i(w); s) = \delta(u, u_i(w)), \quad (3.23)$$

where  $\delta$  is the delta distribution. We multiply both sides of (3.23) by  $w_3^{-s}$  and take the limit as  $w_3 \rightarrow 0$ . Since  $u$  and  $u_i(w)$  become disjoint, the right side of (3.23) vanishes. From the expansion (3.20), we obtain

$$\begin{aligned} \lim_{w_3 \rightarrow 0} (\Delta_\Gamma - s(2-s))w_3^{-s}G_\Gamma(u; u_i(w); s) \\ = (\Delta_\Gamma - s(2-s))E_i(u; u_i^\infty(w'); s) = 0, \end{aligned}$$

where  $u_i^\infty(w')$  are the boundary coordinate functions. For convenience, we simply write  $w'$  when there is no confusion. The idea of obtaining generalized eigenfunctions as the weighted asymptotic limits of the Green's function is quite general.

(b)  $E_i(u; w'; s)$  is the local expression for a section of a line bundle over  $B = \partial_\infty \mathcal{M}$ . To prove this, we compute the transition function which determines the relation between two local expressions for the section in overlapping coordinate charts on  $B$ . Suppose  $U_i \cap U_j \neq \emptyset$  and that  $u_i(w) = u_j(y)$  for  $w \in U_i$  and  $y \in U_j$ . Then  $\exists \gamma \in \operatorname{Isom} \mathbb{H}^3$  so that  $y = \gamma w$ . As described in §3.2, these neighborhoods extend to neighborhood  $U_i^\infty$  and  $U_j^\infty$  of  $B$  with coordinate functions  $u_i^\infty$  and  $u_j^\infty$  and  $u_j^\infty$ , respectively. Now, suppose  $w'$  and  $y'$  represent the same point of  $B$ , i.e.  $u_i^\infty(w') = u_j^\infty(y')$ . From the comment above,  $y' = \lim_{w_3 \rightarrow 0} (\gamma w)'$ . We compute  $E_i(u; u_i^\infty(w'); s)$  in terms of  $E_j(u; u_j^\infty(y'); s)$  using the definition of  $E_i$  in terms of  $G_\Gamma$  and these coordinate relations.

$$\begin{aligned} E_i(u; u_i^\infty(w'); s) &= \lim_{w_3 \rightarrow 0} w_3^{-s} G_\Gamma(u; u_i(w); s) \\ &= \lim_{w_3 \rightarrow \infty} (w_3^{-s} y_3^s) y_3^{-s} G_\Gamma(u; u_j(y); s) \\ &= \lim_{w_3 \rightarrow 0} \left( \frac{w_3}{(\gamma w)_3} \right)^{-s} ((\gamma w)_3^{-s} G_\Gamma(u; u_j((\gamma w)'), (\gamma w)_3); s) \\ &= \tau_\gamma^s E_j(u; u_j^\infty((\gamma w)'); s) \\ &= \tau_\gamma^s E_j(u; u_j^\infty(y'); s), \end{aligned} \quad (3.24)$$

where

$$\tau_\gamma = \lim_{w_3 \rightarrow 0} \left( \frac{(\gamma w)_3}{w_3} \right).$$



These functions  $\tau_\gamma$  are transition functions for a line bundle  $\mathcal{M}_s$  over  $B$ . The transformation law (3.24) shows that  $E_i$  are the local expressions in  $U_i^\infty$  for a section  $E_\Gamma(u; b; s)$  of this line bundle. We let  $\Gamma(\mathcal{M}_s)$  denote the space of all smooth sections on  $\mathcal{M}_s$  (recall that  $B$  is compact for the present case). The transition function  $\tau_\gamma$  can be computed explicitly. It is nothing but the conformal factor discussed in § 1.6. For

$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{PSL}(2, \mathbb{C})$ , we have

$$\tau_\gamma = |cw' + d|^{-2}.$$

### 3.8 Functional equation for the Green's function

The sections  $E_\Gamma$  can be used to obtain densities on  $B$ . Suppose that  $e(b; s) \in \Gamma(\mathcal{M}_s)$ , the smooth sections over  $B$  transforming locally as in (3.24). There is a natural duality between  $\Gamma(\mathcal{M}_s)$  and  $\Gamma(\mathcal{M}_{2-s})$ . If  $f(b; 2-s) \in \Gamma(\mathcal{M}_{2-s})$  with local expressions  $f_i$ , then the two-form locally defined by

$$e_i(w'; s) f_i(w'; 2-s) dw_1 \wedge dw_2, \quad (3.25)$$

is easily seen to be invariant. Indeed, if  $y' \in U_i^\infty$  and  $w' \in U_i^\infty$  are related by  $y' = \lim_{w_3 \rightarrow 0} (\gamma w)'$ ,  $\gamma \in \text{Isom } \mathbb{H}^3$ , then from (3.24),

$$\begin{aligned} e_i(w'; s) f_i(w'; 2-s) &= \tau_\gamma^s \tau_\gamma^{2-s} e_j(y'; s) f_j(y'; 2-s) \\ &= \tau_\gamma^2 e_j(y'; s) f_j(y'; 2-s). \end{aligned}$$

On the other hand, as in § 1.6, one has

$$dy_1 \wedge dy_2 = \tau_\gamma^2 dw_1 \wedge dw_2,$$

proving the invariance of (3.25). This 2-form can be integrated over  $B$ . We will write

$$\int_B e(w'; s) f(w'; 2-s) dw', \quad (3.26)$$

for the integral of the 2-form defined in (3.25) (see (2.1) for the definition of (3.26)). Let us note that if  $s$  is on the critical line  $\text{Re } s = 1$ , then so is  $2-s$ .

**Theorem 3.6.** *If  $\text{Re } s = 1$ ,  $s \neq 1$ , and  $E_\Gamma(u; w'; s) \in \Gamma(\mathcal{M}_s)$  is the section defined in § 3.7, then*

$$G_\Gamma(u; v; s) - G_\Gamma(u; v; 2-s) = (2-2s) \int_B E_\Gamma(v; w'; s) E_\Gamma(u; w'; 2-s) dw'. \quad (3.27)$$

*Sketch of the Proof*

neighborhood  $\{z \mid \|z\|_E < 1\}$ . We choose  $f$  and  $g$  as functions on  $\mathcal{M}$  supported in a regular neighborhood  $U_i$  and such that  $\text{supp}(fg)(u_i(z)) \cap \Omega_a \subset \{z \mid z_3 = a\}$ . Green's theorem states that

$$\begin{aligned} & \int_{\Omega_a} [(f\Delta_\Gamma g)(u_i(z)) - (g\Delta_\Gamma f)(u_i(z))] d\mu(z) \\ &= a^{-1} \int \left[ f(u_i(z)) \frac{\partial g}{\partial z_3}(u_i(z)) \Big|_{z_3=a} - g(u_i(z)) \frac{\partial f}{\partial z_3}(u_i(z)) \Big|_{z_3=a} \right] dz_1 dz_2. \end{aligned} \quad (3.28)$$

If  $\chi_i$  and  $\chi_j$  have support in regular neighborhoods of infinity, we set

$$\begin{aligned} f(\cdot) &= \chi_i(\cdot) G_\Gamma(u; \cdot; s) \\ g(\cdot) &= \chi_j(\cdot) G_\Gamma(\cdot; v; 2-s), \end{aligned}$$

for  $u, v \in \mathcal{M}$  fixed, and substitute them into (3.28). We then insert the asymptotics for  $G_\Gamma$  and  $\partial G_\Gamma / \partial z_3$  from Theorem 3.5. This gives

$$\begin{aligned} & \int_{\Omega_a} [\chi_i G_\Gamma(u; u_i(z); s) \Delta_\Gamma(\chi_j G(u_i(z); v; 2-s)) \\ & \quad - \chi_j G_\Gamma(u_i(z); v; 2-s) \Delta_\Gamma(\chi_i G(u; u_i(z); s))] d\mu(z) \\ &= a^{-1} \int \left\{ \chi_i(a^s E_\Gamma(u; z'; s) + \mathcal{O}(a^{\text{Re } s+1})) \left[ \frac{\partial \chi_j}{\partial z_3}(a^{2-s} E_\Gamma(v; z'; 2-s) \right. \right. \\ & \quad \left. \left. + \mathcal{O}(a^{3-\text{Re } s}) \right) + \chi_j((2-s)a^{1-s} E_\Gamma(v; z'; 2-s) + \mathcal{O}(a^{2-\text{Re } s})) \right] \\ & \quad \left. - [\text{similar terms}] \right\} dz_1 dz_2. \end{aligned}$$

We now take the limit as  $a \rightarrow 0$ . This gives

$$\begin{aligned} & \int_{\mathcal{M}} \chi_i G_\Gamma(u; w; s) \Delta_\Gamma(\chi_j G(w; v; 2-s)) - \chi_j G_\Gamma(w; v; 2-s) \Delta_\Gamma(\chi_i G(u; w; s)) \\ &= (2-2s) \int_B \tilde{\chi}_i(z') \tilde{\chi}_j(z') E_\Gamma(u; z'; s) E_\Gamma(v; z'; 2-s) dz_1 dz_2. \end{aligned} \quad (3.29)$$

Here we write  $\tilde{\chi}_i(z') \equiv \chi_i(u_i^\infty(z'))$ . Since such functions  $\tilde{\chi}_i$  form a covering of  $B$ , we obtain the result from (3.29) by summing over all  $i$  and  $j$  and using the equation satisfied by  $G_\Gamma$ .  $\square$

### 3.9 Eigenfunction expansion and diagonalization

We now come to the main result on the spectral fine structure. We show that the generalized eigenfunctions  $E_\Gamma(u; b; s)$  provide an eigenfunction expansion for  $\mathcal{H}_{\text{ac}}(\Delta_\Gamma)$ , the absolutely continuous subspace of  $L^2(\mathcal{M})$  for  $\Delta_\Gamma$ . These eigenfunctions also provide a spectral representation for the absolutely continuous part of  $\Delta_\Gamma$ , in which it is diagonal. Let  $E_{\text{ac}}(\Delta_\Gamma)$  be the projection of  $L^2(\mathcal{M})$  onto  $\mathcal{H}_{\text{ac}}(\Delta_\Gamma)$ .

Recall that for  $s \in \mathbb{C}$ ,  $\mathcal{M}_s$  denotes the complex line bundle over  $B$  with transition function  $\tau_\gamma^s$ . We denote by  $\Gamma(\mathcal{M}_s)$  the space of smooth sections on this bundle. The space  $\Gamma(\mathcal{M}_{2-s})$  is naturally dual to  $\Gamma(\mathcal{M}_s)$  in that  $e(b; s)e(b; 2-s)db$  is an invariant 2-form on  $B$ . This was verified in § 3.8. We are interested in the critical line  $\operatorname{Re} s = 1$ , so we now write

$$s = 1 + ik, \quad k \in \mathbb{R},$$

and note that  $2 - s = 1 - ik$ . We can identify  $e(b; 1 + ik)^*$  with the section  $e(b; 1 - ik)$  and hence identify  $\Gamma(\mathcal{M}_{1+ik})$  with a dense subspace of a complex Hilbert space as follows. For sections  $e, f \in \Gamma(\mathcal{M}_{1+ik})$ , we define an inner product

$$\langle e, f \rangle = \int_B e(b; 1 + ik) f(b; 1 + ik)^* db. \quad (3.30)$$

Similarly, we can identify  $\Gamma(\mathcal{M}_{1-ik})$  with a Hilbert space. We denote by  $\Gamma_2(\mathcal{M}_{1 \pm ik})$  the two Hilbert spaces obtained from the completion of  $\Gamma(\mathcal{M}_{1 \pm ik})$  in the inner product (3.30) (with appropriate signs in (3.30)).

**Theorem 3.7.** *The maps*

$$\mathcal{F}_\pm : E_{ac}(\Delta_\Gamma) L^2(\mathcal{M}) \rightarrow \int_{(0, \infty)}^\oplus \Gamma_2(\mathcal{M}_{1 \pm ik}) dk$$

defined on  $\mathcal{H}_{ac}(\Delta_\Gamma)$  by

$$(\mathcal{F}_\pm \psi)(k, w') \equiv \left(\frac{2}{\pi}\right)^{1/2} k \int_{\mathcal{M}} E_\Gamma(u; w'; 1 \pm k) \psi(u) d\mu(u) \quad (3.31)$$

are isometries which diagonalize  $\Delta_\Gamma$ . For  $\psi \in D(\Delta_\Gamma) \cap \mathcal{H}_{ac}(\Delta_\Gamma)$ ,

$$(\mathcal{F}_\pm \Delta_\Gamma \psi)(k, w') = (1 + k^2)(\mathcal{F}_\pm \psi)(k, w').$$

Note the similarity between the maps  $\mathcal{F}_\pm$  defined in (3.31) for  $\mathcal{M}$  and those in (2.13), which give the spectral representation for  $\Delta_E$  on  $\mathbb{R}^n$ , and those in (2.42), which give the spectral representation for  $\Delta_{\mathbb{H}^n}$  on  $L^2(\mathbb{H}^n)$ . In all three cases, the kernel of the transformation is given by the generalized eigenfunctions obtained from certain weighted asymptotic limits of the Green's function.

*Sketch of the Proof*

We show here only that for  $\psi \in \mathcal{H}_{ac}(\Delta_\Gamma)$ ,

$$\|\mathcal{F}_\pm \psi\| = \|E_{ac}(\Delta_\Gamma) \psi\|_{L^2(\mathcal{M}, d\mu)}, \quad (3.32)$$

(one also has to show that  $\operatorname{Ran} \mathcal{F}_\pm$  are dense; see [FHP2]). For  $1 < a < b < \infty$  (the  $(1 + k^2)$ -values), set  $a' = (a - 1)^{1/2}$  and  $b' = (b - 1)^{1/2}$  (the corresponding  $k$  parameters). To verify (3.32), it suffices to show that

$$\int_{a'}^{b'} \|(\mathcal{F}_\pm \psi)_k\|_{\Gamma_2(\mathcal{M}_{1 \pm ik})}^2 dk = \|E_{(a,b)}(\Delta_\Gamma) \psi\|_{L^2(\mathcal{M}, d\mu)}^2. \quad (3.33)$$

We calculate the left side of (3.33),

$$\begin{aligned}
 & \int_{a'}^{b'} \|(\mathcal{F} \pm \psi)_k\|_{\Gamma_{2l, \mathcal{H}_1 + ik}} dk \\
 &= \int_{a'}^{b'} dk \int_B \left(\frac{2}{\pi}\right) k^2 \left\{ \int_M E_\Gamma(u; w'; 1 \mp ik) \psi^*(u) d\mu(u) \right\} \\
 & \qquad \qquad \qquad \left\{ \int_{\mathcal{H}} E_\Gamma(v; w'; 1 \pm ik) \psi(v) d\mu(v) \right\} dw' \\
 &= \int_{a'}^{b'} dk \left(\frac{2}{\pi}\right) k^2 \int_{\mathcal{H}} d\mu(u) \int_{\mathcal{H}} d\mu(v) \psi^*(u) \psi(v) \\
 & \qquad \qquad \qquad \left\{ \int_B dw' E_\Gamma(u; w'; 1 \mp ik) E_\Gamma(v; w'; 1 \pm ik) \right\}.
 \end{aligned} \tag{3.34}$$

The last factor can be rewritten using the functional equation for  $G_\Gamma$ , (3.27),

$$\begin{aligned}
 & \int_B E_\Gamma(u; w'; 1 \mp k) E_\Gamma(v; w'; 1 \pm k) db \\
 &= \pm \frac{i}{2k} [G_\Gamma(u; v; 1 \pm ik) - G_\Gamma(u; v; 1 \mp ik)].
 \end{aligned}$$

Inserting this into (3.34), we obtain for each  $k$ ,

$$\|(\mathcal{F} \pm \psi)_k\|^2 = \pm \frac{ik}{\pi} \int_{\mathcal{H} \times \mathcal{H}} \psi^*(u) [G_\Gamma(u; v; 1 \pm ik) - G_\Gamma(u; v; 1 \mp ik)] \psi(v). \tag{3.35}$$

We now apply Stone's formula in  $\mathcal{H}_{ac}(\Delta_\Gamma)$ ,

$$E_{(a,b)}(\Delta_\Gamma) = \pm \frac{1}{\pi i} \int_{a'}^{b'} k [G_\Gamma(k) - G_\Gamma(-k)] dk,$$

where  $G_\Gamma(k) = G_\Gamma(1 \pm ik)$ , etc., to obtain from (3.35),

$$\int_{a'}^{b'} \|(\mathcal{F} \pm \psi)_k\|_{\Gamma^2(\mathcal{H}_1 \pm ik)} dk = \|E_{(a,b)}(\Delta_\Gamma) \psi\|_{L^2(\mathcal{H}, d\mu)}^2$$

proving (3.33). □

#### 4. Eisenstein series and the $S$ -matrix

In this chapter, we sketch some further developments in the spectral theory of non-compact hyperbolic manifolds. The eigenfunction expansion and spectral representation obtained in chapter 3 is but a necessary prerequisite for the study of

$\Gamma < \text{Isom } \mathbb{H}^n$ , the  $S$ -matrix, the Selberg zeta function, and trace formulas. We will discuss the first two of these topics here. We will conclude in Chapter 5 with a brief overview of the other two topics and the question of embedded eigenvalues. For other approaches to the  $S$ -matrix and the Eisenstein series for the case when  $\Gamma$  has no parabolic elements, see [Ma], [Pa2], and [Pe3]. The results in this section are discussed in detail in [FHP2] and [FHP3].

#### 4.1 Classical Eisenstein series

To construct the classical Eisenstein series associated with a discrete geometrically finite subgroup  $\Gamma < \text{Isom } \mathbb{H}^n$ , we begin with the simplest case  $\Gamma = \{id\}$ , the trivial subgroup. For  $z \in \mathbb{H}^n$ ,  $b \in \partial_\infty \mathbb{H}^n \cong \mathbb{R}^{n-1}$ , and  $s \in \mathbb{C}$  with  $\text{Re } s > n - 1$ , we define a function  $\tilde{E}_0(z; b; s)$  as follows:

$$\begin{aligned} b = \infty \quad \tilde{E}_0(z; b; s) &= z_n^s \\ b \in \mathbb{R}^{n-1} \quad \tilde{E}_0(z; b; s) &= \left( \frac{z_n}{\|z' - b\|_E^2 + z_n^2} \right)^s, \end{aligned} \quad (4.1)$$

where we write  $z = (z', z_n) \in \mathbb{R}^{n-1} \times \mathbb{R}^+$ , as above. The function  $\tilde{E}_0$  has the following properties,

- (i)  $\Delta_{\mathbb{H}^n, z} \tilde{E}_0(z; b; s) = s(n - 1 - s) \tilde{E}_0(z; b; s)$ ;
- (ii)  $\tilde{E}_0$  is smooth in all variables;
- (iii) for  $z' \neq b$ , and  $\text{Re } s > (n - 1)/2$ ,

$$\int_0^1 \frac{dz_n}{z_n} |\tilde{E}_0(z; b; s)|^2 < \infty,$$

but the integral diverges for  $\text{Re } s \leq (n - 1)/2$ . These generalized eigenfunctions  $\tilde{E}_0$  are related to those constructed in Example 2.11, (2.39).

Turning now to the subgroup  $\Gamma$ , we formally construct a  $\Gamma$ -automorphic function  $\tilde{E}_\Gamma(z; b; s)$  from  $\tilde{E}_0$  by periodizing  $\tilde{E}_0$  with respect to  $\Gamma$ . As above, we suppose that  $\Gamma$  has no elliptic elements which implies that the stabilizer subgroup of any  $z \in \mathbb{H}^n$  in  $\Gamma$  is trivial. We define  $\tilde{E}_\Gamma$  as a series

$$\tilde{E}_\Gamma(z; b; s) \equiv \sum_{\gamma \in \Gamma} \tilde{E}_0(\gamma z; b; s). \quad (4.2)$$

Since  $\tilde{E}_\Gamma$  is  $\Gamma$ -automorphic,  $\tilde{E}_\Gamma \in D(\Delta_\Gamma)$ . Moreover,  $\tilde{E}_\Gamma$  satisfies

$$\Delta_\Gamma \tilde{E}_\Gamma(z; b; s) = s(n - 1 - s) \tilde{E}_\Gamma(z; b; s).$$

As for the convergence of the sum (4.2), we have

*Lemma 4.1. The sum (4.2) converges uniformly for  $\text{Re } s > n - 1$  and in this half-plane,  $\tilde{E}_\Gamma$  is a classical generalized eigenfunction of  $\Delta_\Gamma$ . More specifically, there exists a constant  $\delta_\Gamma$ , called the exponent of convergence, satisfying  $(n - 1)/2 < \delta_\Gamma \leq n - 1$ , so that the sum (4.2) converges for  $\text{Re } s > \delta_\Gamma$ .*

The functions  $E_\Gamma$  are called the *Eisenstein series* for  $\Gamma$ . For a proof of Lemma 4.1, see, for example, [A1].

#### 4.2 Connection between $\tilde{E}_\Gamma$ and the generalized eigenfunctions $E_\Gamma$

The Eisenstein series  $\tilde{E}_\Gamma$  of §4.1 are related to the generalized eigenfunctions  $E_\Gamma$  of §3.7 as follows. Let  $G_0(u; v; s)$  be the Green's function for  $\mathbb{H}^n$  (see section 3.5). We can periodize  $G_0$  as in (4.2) above and obtain a  $\Gamma$ -automorphic function  $G_\Gamma$ ,

$$G_\Gamma(u; v; s) \equiv \sum_{\gamma \in \Gamma} G_0(\gamma u; v; s). \quad (4.3)$$

One can show as in §4.1 that this sum converges for  $\operatorname{Re} s > n - 1$  (more specifically, for  $\operatorname{Re} s > \delta_\Gamma$ ). The function  $G_\Gamma$  satisfies

$$(\Delta_\Gamma - s(n - 1 - s))G_\Gamma(u; v; s) = \delta(u, v)$$

for  $u, v \in F_\Gamma$ , a fundamental domain for  $\Gamma$ . Hence,  $G_\Gamma$  is the Green's function for  $\Delta_\Gamma$ . Let us use the representation (4.3) to construct  $E_\Gamma$  by the method of §3.7. We write  $v = (b, v_n)$ , be  $\mathbb{R}^{n-1}$ . From Lemma 3.3, we have the asymptotic formula,

$$G_0(u; (b, v_n); s) = v_n^s u_n^s (g_{b,s}(u) + \mathcal{O}(v_n)).$$

Inserting this asymptotic expression into (4.3) and taking the limit as  $v_n \rightarrow 0$ , we obtain for  $\operatorname{Re} s > n - 1$ ,

$$\begin{aligned} \lim_{v_n \rightarrow 0} v_n^{-s} G_\Gamma(u; (b, v_n); s) &= \sum_{\gamma \in \Gamma} (\gamma u)_n^s g_{b,s}(\gamma u) \\ &= \sum_{\gamma \in \Gamma} E_0(\gamma u; b; s) \\ &= \tilde{E}_\Gamma(u; b; s), \end{aligned} \quad (4.4)$$

where we used the definition of  $E_0$ , the equality of  $E_0$  and  $\tilde{E}_0$ , and the definition of  $\tilde{E}_\Gamma$ . Since the left side of (4.4) converges to  $E_\Gamma(u; b; s)$  for  $\operatorname{Re} s > (n - 1)/2$ ,  $s(n - 1 - s) \notin \sigma_d(\Delta_\Gamma)$ , we obtain the identity,

$$E_\Gamma(u; b; s) = \tilde{E}_\Gamma(u; b; s), \quad (4.5)$$

for  $\operatorname{Re} s > n - 1$ , the half-plane on which  $\tilde{E}_\Gamma$  is defined. This identifies the generalized eigenfunctions  $E_\Gamma$  of  $\Delta_\Gamma$  with the classical Eisenstein series for  $\Gamma$ , when  $\operatorname{Re} s > n - 1$ . Moreover, (4.4)–(4.5) show that the eigenfunctions  $E_\Gamma$  provide a meromorphic continuation of the Eisenstein series to  $\operatorname{Re} s > (n - 1)/2$  and a continuous extension onto the critical line  $\operatorname{Re} s = (n - 1)/2$ ,  $s \neq (n - 1)/2$  or an embedded eigenvalue of  $\Delta_\Gamma$ .

One of the main motivation for our work is the following question:

Do the Eisenstein series for  $\Gamma$ , a geometrically finite, torsion-free, discrete subgroup of  $\operatorname{Isom} \mathbb{H}^n$ , admit a meromorphic continuation to the entire complex plane?

The answer is yes. We will sketch the ideas behind the proof of this statement in the next two sections.

### 4.3 The $S$ -matrix

For fixed  $u \in \mathcal{M}$ , the Eisenstein series  $E_\Gamma(u; b; s)$  are smooth sections on  $B = \partial_\infty \mathcal{M}$ . Let us assume that  $\Gamma$  is purely hyperbolic. As we have seen,  $B$  is then a smooth compact manifold. In § 3.8, we constructed a Hilbert space  $\Gamma_2(\mathcal{M}_s)$ ,  $s = 1 + ik$ ,  $k \in \mathbb{R}$ , of sections of the complex line bundle  $\mathcal{M}_s$  over  $B$ . The  $S$ -matrix associated with  $\Gamma$  is a unitary operator  $S(s)$  initially defined on  $\Gamma_2(\mathcal{M}_{1+ik})$ . As we shall see, its dependency on  $s$  is such that it admits a meromorphic continuation in  $s$  to a linear operator acting between certain Banach spaces of sections over  $\mathcal{M}_s$ .

The  $S$ -matrix can be introduced by means of the functional equation for the Green's function (3.27). For  $\operatorname{Re} s = 1$ ,  $s \neq 1$ , we have

$$G_\Gamma(u; v; s) = G_\Gamma(u; v; 2-s) + (2-2s) \int_B E_\Gamma(v; w'; s) E_\Gamma(u; w'; 2-s). \quad (4.6)$$

We now multiply both sides of (4.6) by  $v_3^{-s}$  and take the limit as  $v_3 \rightarrow 0$ . The limit is evaluated by means of the asymptotics for  $G_\Gamma$  given in Theorem 3.5 and the following result on the asymptotics of the Eisenstein series.

**Theorem 4.2.** *The Eisenstein series  $E_\Gamma(u; b; s)$ , for  $\operatorname{Re} s \geq 1$ ,  $s \neq 1$  and  $(2-s)s \notin \sigma_d(\Delta_\Gamma)$ , can be written in a regular neighborhood  $U_i$  as*

$$E_{\Gamma,i}(u_i(x); b; s) = E_0(x; b; s) + K_i(x; b; s), \quad (4.7)$$

where  $E_0$  is given by

$$E_0(x; b; s) = \frac{x_3}{(2\pi)^2 2^{s-1} \Gamma(s)} \int_{\mathbb{R}^2} e^{i\xi(x'-b)} |\xi|^{s-1} K_{s-1}(x_3 |\xi|) d^2 \xi,$$

(see (2.39)) and equals  $E_0$  in (4.1), and the remainder  $K_i$  is given by

$$K_i(x; b; s) = x_3^s k_i(x'; b; s) + \mathcal{O}(x_3^{\operatorname{Re} s + 1}) \quad (4.8)$$

for a function  $k_i$  smooth in  $x'$  and  $b$ .

Returning to (4.6), note that if we write

$$v_3^{-s} G_\Gamma(u; v; 2-s) = v_3^{-2} (v_3^{2-s} G_\Gamma(u; v; 2-s)), \quad (4.9)$$

then the term in parentheses converges to  $E_\Gamma(u; v'; 2-s)$  but the coefficient is singular. This singularity is cancelled by the  $E_0(v; w'; s)$ -contribution to the integral in (4.6) coming from  $E_\Gamma$ . Note that the error  $K_i$  in (4.7) does not contribute to the asymptotic limit as is clear from (4.8). The result of this calculation indicates that the right side of (4.6) can be expressed as the action of a linear operator on  $E_\Gamma(u; \cdot; 2-s)$ :

$$\begin{aligned} \lim_{v_3 \rightarrow \infty} \left\{ v_3^{-s} G_\Gamma(u; v; 2-s) + (2-2s) \int_B v_3^{-s} E_\Gamma(v; w'; s) E_\Gamma(u; w'; 2-s) \right\} \\ \equiv S(s) E(u; \cdot; 2-s). \end{aligned}$$

As the left side of (4.6) multiplied by  $v_3^{-s}$  converges to  $E_\Gamma(u; v'; s)$ , we obtain the important identity for  $\operatorname{Re} s = 1$ ,  $s \neq 1$ ,

$$E_\Gamma(u; v'; s) = S(s) E_\Gamma(u; \cdot; 2-s). \quad (4.10)$$

The linear operator  $S(s)$  on  $\Gamma_2(\mathcal{M}_s)$ , called the  $S$ -matrix, is actually unitary for  $\operatorname{Re} s = 1$ . On the right side of the critical line,  $\operatorname{Re} s > 1$ ,  $S(s)$  is a linear map from  $\Gamma(\mathcal{M}_{(n-1)/2-s}) \rightarrow \Gamma(\mathcal{M}_s)$ . When  $\Gamma$  is the trivial group, we can compute the  $S$ -matrix explicitly from the results of Example 2.11 using the standard asymptotics of the modified Bessel function. Since  $\partial_\infty \mathbb{H}^n = \widehat{\mathbb{R}}^{n-1} = S^n$ , we can express its action in local (flat) coordinates. In the Fourier transform variable  $\xi'$  of Example 2.11, for  $f \in C_0^\infty(\mathbb{R}^{n-1})$ , we have

$$(S(1+ik)f)^\wedge(\xi') = c(s) \|\xi'\|_E^{2ik} \widehat{f}(\xi'), \quad (4.11)$$

where  $\wedge$  denotes the Fourier transform,  $s = (n-1)/2 + ik$ , with  $k \in \mathbb{R}$  and  $c(s)$  is defined by

$$c(s) = 2^{-2ik} \frac{\Gamma(1-ik)}{\Gamma(1+ik)}.$$

This shows that  $S$  acts locally as a complex power of the Laplacian on  $\mathbb{R}^{n-1}$ ,

$$S(s) = c(s) \Delta_B^{s-(n-1)/2}. \quad (4.12)$$

Such a representation is valid only locally. There is no *a priori* metric on  $B$  and hence no distinguished Laplacian. However,  $S(s)$  is an elliptic pseudo-differential operator and has a local kernel of the type indicated by expressions (4.11)–(4.12). For points  $b$  and  $c$  belonging to a coordinate patch of  $B$ , the principal part of the kernel is given by

$$S(b; c; s) = c(s) \|b - c\|_E^{-2s},$$

up to smooth contributions. In general, we have the

**Theorem 4.3.** *Let  $\Gamma$  be discrete and purely hyperbolic so that  $B = \partial_\infty \mathcal{M}$  is a compact, connected manifold. The  $S$ -matrix  $S(s)$  is an elliptic pseudo-differential operator with principal symbol*

$$c(s) \|\xi\|_E^{2s-n+1}$$

*acting between  $\Gamma(\mathcal{M}_{n-s-1}) \rightarrow \Gamma(\mathcal{M}_s)$ , meromorphic for  $\operatorname{Re} s > (n-1)/2$  and unitary on  $\operatorname{Re} s = (n-1)/2$ ,  $s \neq (n-1)/2$ .*

We mention that when  $\operatorname{Vol} \mathcal{M} < \infty$  and there are MR cusps, the  $S$ -matrix is, in fact, a matrix, the size of which is determined by the number of cusps. When there are NMR cusps, the  $S$ -matrix is still a pseudo-differential operator but it has precise asymptotics which reflect the non-compact nature of  $B$ .

#### 4.4 Meromorphic continuation of the Eisenstein series

We can now present one of the main results of the theory developed here: the meromorphic continuation of the Eisenstein series.

**Theorem 4.4.** *Let  $\Gamma$  be a discrete, torsion-free, geometrically finite subgroup of hyper-*



bolic isometries such that  $\mathcal{M} \equiv \mathbb{H}^n/\Gamma$  is non-compact. Let  $E_\Gamma(u; b; s)$  be the corresponding Eisenstein series as defined in (4.1) with  $\operatorname{Re} s > (n-1)$ . Then  $E_\Gamma$  extends to a meromorphic function on  $\mathbb{C} \setminus \{(n-1)/2\}$ .

### *Sketch of the Proof*

As above, let us assume  $\Gamma$  is purely hyperbolic so that  $B$  is compact. In this case,  $S(s)$  is an elliptic pseudo-differential operator on a compact manifold according to Theorem 4.2. We begin with the functional equation (4.10) for  $\operatorname{Re} s = (n-1)/2$ ,  $s \neq (n-1)/2$ :

$$E_\Gamma(u; b; s) = S(s) E_\Gamma(u; \cdot; (n-1-s)).$$

On the critical line,  $S(s)$  is unitary and hence invertible. We write the functional equation as

$$S(s)^{-1} E_\Gamma(u; \cdot; s) = E_\Gamma(u; b; (n-1-s)). \quad (4.13)$$

The strategy is as follows. On the right side of (4.13), the Eisenstein series  $E_\Gamma(u; b; n-1-s)$  is meromorphic by Theorem 4.1, for  $\operatorname{Re} s < (n-1)/2$ . We wish to extend this to  $\operatorname{Re} s \geq (n-1)/2$ . On the left side of (4.13), the Eisenstein series  $E_\Gamma(u; b; s)$  is meromorphic for  $\operatorname{Re} s > (n-1)/2$  and continuous on  $\operatorname{Re} s = (n-1)/2$ . Hence, it remains to prove that  $S(s)^{-1}$  exists and is meromorphic for  $\operatorname{Re} s > (n-1)/2$ . This follows from an explicit construction of a paramatrix  $R(s)$  for  $S(s)$  satisfying

$$R(s)S(s) = 1 + K(s), \quad (4.14)$$

where  $K(s)$  is a meromorphic, compact operator-valued function on  $\operatorname{Re} s > (n-1)/2$ . The paramatrix  $R(s)$  is also meromorphic in  $\operatorname{Re} s > (n-1)/2$ . The operator  $1 + K(s)$  is invertible in  $\operatorname{Re} s > (n-1)/2$  by the meromorphic Fredholm theorem. Hence, we obtain an inverse,

$$S(s)^{-1} = (1 + K(s))^{-1} R(s),$$

for  $\operatorname{Re} s > (n-1)/2$ . This continuation of  $S(s)^{-1}$  can be controlled well enough to allow one to prove that the left side of (4.13) has a continuation into  $\operatorname{Re} s > (n-1)/2$ . This establishes the continuation of the right side of (4.13) and, hence, of the Eisenstein series.  $\square$

## 5. Remarks on current research directions

We complete this survey with a brief description of several open problems in the area and references to some recent work. As mentioned in the introduction, the spectral theory developed here is only a necessary prerequisite for the exploration of the interesting relationships between the geometry and analysis of hyperbolic manifolds.

hyperbolic 3-manifolds, there are some very interesting results concerning the bottom of the spectrum; see the article of Canary [C] for a discussion.

ii. *Asymptotically constant negative curvature manifolds.* Many of the results on the asymptotic behavior of the Green's function and the behavior of the generalized eigenfunctions can be obtained for non-compact Riemannian manifolds for which the metric approaches the hyperbolic metric at infinity and for which the boundary at infinity is a smooth, compact manifold. The works of Mazzeo and Melrose [MM] and of Agmon [Ag3] treat this situation. The paper of Agmon [Ag3] presents a beautiful representation theorem for solutions of  $\Delta_\Gamma u = \lambda u$ , when  $\Gamma$  has no parabolics, in terms of hyperfunctions on the boundary at infinity. One can also study asymptotically constant negative curvature manifolds as perturbations of hyperbolic manifolds using the techniques of scattering theory; see Perry [Pe2], Guillopé [Gu], and DeBièvre-Hislop-Sigal [DeBHS].

iii. *Selberg zeta function.* Let  $\Gamma < \text{Isom } \mathbb{H}^n$  be purely hyperbolic. Any  $\gamma \in \Gamma$  fixes two points on  $\partial_\infty \mathbb{H}^n$  and the corresponding geodesic joining them. There exists a hyperbolic isometry which maps these two points to 0 and  $\infty$ , respectively. Then  $\gamma$  is conjugate to a dilation by  $\lambda$  (see Example 1.20) and a rotation about the half-line  $x_n \in (0, \infty)$  in  $\mathbb{H}^n$ . Let  $l(\gamma) = |\log \lambda|$ , the length of the minimal closed geodesic associated with the dilation (see the discussion of the Selberg trace formula in section 2.2). Let  $\{\alpha_i\}_{i=1}^{n-1}$  be the eigenvalues of the rotation matrix. If  $\gamma$  is a primitive hyperbolic element, we associate a function  $Z_\gamma(s)$  with  $\gamma$  by

$$Z_\gamma(s) \equiv \prod_{k_i \in \mathbb{N} \cup \{0\}} (1 - \alpha_1^{k_1} \alpha_2^{k_2} \dots \alpha_{n-1}^{k_{n-1}} e^{-(s+k_1+\dots+k_{n-1})l(\gamma)})^2.$$

The Selberg zeta function for  $\Gamma$  is defined as the product of such  $Z_\gamma$ 's over all inconjugate primitive (i.p.) elements

$$Z_\Gamma(s) = \prod_{\text{i.p. } \gamma \in \Gamma} Z_\gamma(s).$$

Using the fact that the number of closed geodesics of length  $\leq l$  grows like  $e^{(n-1)l}$ , one can prove the convergence of the products for  $\text{Re } s > (n-1)$ . For a variety of reasons (see (iv) below), one would like to know that  $Z_\Gamma$  has a meromorphic extension to  $\mathbb{C}$ . This has only recently been proved by Patterson and Perry [PaPe]. These authors relate the poles of the logarithmic derivative of the zeta function to the poles of the  $S$ -matrix and eigenvalues of the Laplacian. The compact case is described in the review of McKean [McK]. It is clear from the construction that  $Z_\Gamma$  encodes geometric information about the manifold  $\mathbb{H}^n/\Gamma$ .

iv. *Trace formulas.* The Selberg trace formula for compact quotients was described in §2.2. Much effort has been recently directed to finding an analogy for the case of non-compact quotients of the type described in this article. The finite volume case has been studied by Lax and Phillips [LP5] and Hedjál [He], among others. In the case of infinite volume with no cusps, Perry [Pe4] has obtained a local trace formula. In analogy with trace formulas occurring in quantum mechanical scattering theory, the trace of the heat kernel in the classical formula is replaced by a renormalized trace of the logarithmic derivative of the  $S$ -matrix. This quantity carries the information about the absolutely continuous spectrum of  $\Delta_\Gamma$ . The right side of the trace formula, carrying the geometric information, is expressed in terms of the Selberg zeta function.

compact case,

$$Z_{\Gamma}(s)Z_{\Gamma}(s)^{-1} = (2s - n + 1) \int_{\mathcal{M}} [G_{\Gamma}(u; u; s) - G_0(u; u; s)] d\mu(u), \quad (5.1)$$

where  $G_0$  is the Green's function on  $\mathbb{H}^n$ . Using asymptotics similar to those proved in this article, Parry re-expressed the right side of (5.1) in terms of the  $S$ -matrix. Such a formula relates the poles and zeros of the meromorphic continuation of the zeta function to those of the  $S$ -matrix. An extension of this result to the general geometrically finite case is presently being pursued.

v. *Resonances*. The estimation of the number of poles of the meromorphic continuation of the  $S$ -matrix for Schrödinger operators and hyperbolic manifolds is a topic of much current research. Such poles are also called resonances. We refer to the recent review article by Zworski [Z] for a comprehensive survey. With regard to hyperbolic manifolds, recent advances have been made by Perry [Pe5], and by Guillopé and Zworski [GZ1, 2]. The problem is as follows. The  $S$ -matrix, as defined in §4.3, admits a meromorphic continuation to  $\operatorname{Re} s < (n - 1)/2$ . Let  $N(r)$  be the number of poles of this continuation in the region  $\{s \in \mathbb{C} \mid \operatorname{Re} s < (n - 1)/2 \text{ and } |s - (n - 1)/2| < r\}$ . We wish to determine upper and lower bounds (a more difficult problem) on  $N(r)$  as  $r \rightarrow \infty$ . (Note that the number of poles in  $\operatorname{Re} s > (n - 1)/2$  is finite if  $\operatorname{Vol} \mathcal{M} = \infty$ ). One does not usually work with the  $S$ -matrix directly but with the Green's function or the zeta function. This problem is the analog of the classical eigenvalue counting problem for compact manifolds (see [McK]). It is known in the compact case that the number of eigenvalues of  $\Delta_{\Gamma}$  less than  $r$  grows like  $r^n$ . Guillopé and Zworski [GZ2] have recently established the bound

$$N(r) \leq Cr^{n+1},$$

for a class of asymptotically constant negative curvature manifolds.

vi. *Embedded eigenvalues*. The existence of embedded eigenvalues for finite volume quotients in 2-dimensions is believed to occur only if  $\Gamma$  is arithmetic (see [DIPS]). Much work (cf. [PS1] and [CdV]) has been devoted to this question. Due to the apparent instability of embedded eigenvalues, various authors (cf. [Mu] and [PS2]) have turned to looking at the resonance set for  $\Gamma$ . This set, roughly speaking, is the union of eigenvalues and resonances (i.e. poles of the  $S$ -matrix) for  $\Gamma$ . It is believed that this set, which is rather stable under perturbations, is a good candidate for the investigation of 'iso-resonance' problems associated with hyperbolic manifolds and their perturbations (recall comment 3 in §2.2).

## References

- [Ag1] Agmon S, Lectures on exponential decay of solutions of second order elliptic equations: Bounds on eigenfunctions of  $N$ -body Schrödinger operators, (1982) (New Jersey: Princeton University Press)
- [Ag2] Agmon S, On the spectral theory of the Laplacian on non-compact hyperbolic manifolds, Journ. "Equations Deriv. Partielles", St. Jean de Monts, 1987, Exp. No. XVII, Palaiseau: École Polytechnique, 1987

- [Ag3] Agmon S, A representation theorem for solutions of Schrödinger type equations on non-compact Riemannian manifolds, *Astérisque* **210** (1992) 13–26
- [A1] Ahlfors L, *Möbius transformations in several dimensions*, University of Minnesota Lecture Notes, 1981
- [B] Beardon A F, *The geometry of discrete groups* (1983) (New York: Springer-Verlag)
- [BeP] Benedetti R and Petronio C, *Lectures on hyperbolic geometry* (1991) (Berlin: Springer-Verlag)
- [Bu] Buser P, Riemannsche flächen mit eigenwerten in  $(0, 1/4)$ , *Comment. Math. Helv.* **52** (1977) 25–34
- [C] Canary R, The Laplacian and the geometry of hyperbolic 3-manifolds, (1990) *Proc. AMS Summer Workshop* (to appear)
- [Ca] do Carmo M P, *Riemannian geometry* (1992) (Boston: Birkhauser)
- [Chr] Charlap L S, *Bierberbach groups and flat manifolds* (1986) (New York: Springer-Verlag)
- [Ch] Chavel I, *Eigenvalues in Riemannian geometry* (1984) (New York: Academic Press)
- [CdV] Colin de Verdière Y, Pseudo-Laplacians II, *Ann. Inst. Fourier* **33** (1983) 87–113
- [CFKS] Cycon H, Froese R, Kirsch W and Simon B, Schrödinger operators, with applications to quantum mechanics and global geometry, *New York: Springer Texts and Monographs in Physics* (1987) (Springer-Verlag)
- [DeBHS] DeBièvre S, Hislop P D and Sigal I M, Scattering theory for the wave equation on non-compact manifolds, *Rev. Math. Phys.* **4** (1992) 575–618
- [DIPS] Deshouillers J M, Iwaniec H, Phillips R S and Sarnak P, Maass cusps forms, *Proc. Natl. Acad. Sci.* **82** (1985) 3533–3534
- [Ep] Epstein C L, The Spectral theory of geometrically periodic hyperbolic 3-manifolds, *Mem. Am. Math. Soc.* (1985) (Providence, R.I.: American Mathematical Society) **58**
- [FHe1] Froese R and Herbst I, Exponential bounds and absence of positive eigenvalues for  $N$ -body Schrödinger operators, *Commun. Math. Phys.* **87** (1982) 429–447
- [FHe2] Froese R and Herbst I, A new proof of the Mourre estimate, *Duke Math. J.* **49** (1982) 4
- [FH] Froese R and Hislop P D, Spectral analysis of second order elliptic operators on non-compact manifolds, *Duke Math. J.* **58** (1989) 103–129
- [FHP1] Froese R, Hislop P D and Perry P, A Mourre estimate and related bounds for the Laplace operator on a hyperbolic manifold with cusps of non-maximal rank, *J. Funct. Anal.* **98** (1991) 292–310
- [FHP2] Froese R, Hislop P D and Perry P, The Laplace operator on hyperbolic manifolds with cusps on non-maximal rank, *Invent. Math.* **106** (1991) 295–333
- [FHP3] Froese R, Hislop P D and Perry P, The Laplace operator on hyperbolic manifolds with irrational cusps, (in preparation)
- [FS] Froese R and Sigal I M, *Lectures in scattering theory* (University of British Columbia) preprint 1992
- [G] Gaffney M P, The Harmonic operator for exterior differential forms, *Proc. Natl. Acad. Sci. USA* **37** (1951) 48–50
- [GWW] Gordon C, Webb D and Wolpert S, Isospectral plane domains and surfaces via Riemannian orbifolds, *Invent. Math.* **110** (1992) 1–22
- [Gu] Guillope L, Théorie spectrale de quelques variétés à bouts, *Ann. scient. Ec. Norm. Sup.* **22** (1989) 137–160
- [GZ1] Guillope L and Zworski M, Upper bounds on the number of resonances for non-compact Riemann surfaces, (1993) preprint
- [GZ2] Guillope L and Zworski M, Polynomial bounds on the number of resonances for some complete spaces of constant negative curvature near infinity, (1993) Prepublication de l'Institut Fourier No. 255
- [He] Hejhal D, The Selberg trace formula for PSL  $(2, \mathbb{R})$ , Vol. 2, *Springer Lecture Notes in Math.*, (1981) (New York: Springer-Verlag) Vol. 1001
- [K] Kac M, Can one hear the shape of a drum?, *Am. Math. Mon.* **73** (1966) 1–23
- [LP1] Lax P and Phillips R S, Translation representation for automorphic solutions of the wave equation in non-euclidean spaces, I, *Commun. Pure Appl. Math.* **37** (1984) 303–328
- [LP2] Lax P and Phillips R S, Translation representation for automorphic solutions of the wave equation in non-euclidean spaces, II, *Commun. Pure Appl. Math.* **37** (1984) 779–813
- [LP3] Lax P and Phillips R S, Translation representation for automorphic solutions of the wave equation in non-euclidean spaces, III, *Commun. Pure Appl. Math.* **38** (1985) 179–208

- [LP4] Lax P and Phillips R S, Translation representation for automorphic solutions of the wave equation in non-euclidean spaces, IV. *Commun. Pure Appl. Math.* **45** (1992) 179–201
- [LP5] Lax P and Phillips R S, Scattering theory for automorphic functions, *Ann. Math. Studies* (1976) (Princeton: University Press) **87**
- [Le] Lebedev N N, *Special functions and their application* (1992) (New York: Dover Publishing Co.),
- [Ma] Mandouvaos N, Scattering operator, inner product formula, and “Maass-Selberg” relations for Kleinian groups, *Mem. Am. Math. Soc.* **400** (1989) (Providence, R.I.: American Mathematical Society)
- [M] Mazzeo R, Unique continuation at infinity and embedded eigenvalues for asymptotically hyperbolic manifolds, *Am. J. Math.* **113** (1991) 25–46
- [MM] Mazzeo R and Melrose R, Meromorphic extension of the resolvent on complete spaces with asymptotically constant negative curvature, *J. Funct. Anal.* **75** (1987) 260–310
- [MP] Mazzeo R and Phillips R S, Hodge theory on hyperbolic manifolds, *Duke Math. J.* **60** (1990) 509–559
- [McK] McKean H, Selberg’s trace formula applied to a compact Riemann surface, *Comm. Pure Appl. Math.* **25** (1975) 225–246
- [Mo] Mourre E, Absence of singular continuous spectrum for certain self-adjoint operators, *Commun. Math. Phys.* **78** (1981) 391–400
- [Mu] Müller W, Spectral geometry and scattering theory for certain complete surface of finite volume, *Invent. Math.* **109** (1992) 265–305
- [Pa1] Patterson S J, Lectures on measures on limits sets of Kleinian groups, in *Analytical and geometric aspects of hyperbolic space*, (ed.) D B A Epstein, (1987) (Cambridge: Cambridge University Press)
- [Pa2] Patterson S J, The Laplacian operator on a Riemann surface, *Compos. Math.* **31** (1975) 83–107; **32** (1976) 71–112, 227–259
- [Pa3] Patterson S J, The Selberg zeta function of a Kleinian group, in *Number theory, trace formulas, and discrete groups*, (eds) Aubert K E, Bombieri E and Goldfeld D (1989) (Boston: Academic Press)
- [PaPe] Patterson S J and Perry P, Divisors of the Selberg zeta function and Kleinian groups (1994) (University of Kentucky) preprint
- [Pe1] Perry P, Inverse spectral problems on compact Riemannian manifolds, in *Schrödinger operators: Proceedings of the Nordic Summer School in Mathematics* (eds) H Holden and A Jensen (1989) (Berlin: Springer-Verlag)
- [Pe2] Perry P, The Laplace operator on a hyperbolic manifold, I. Spectral and scattering theory, *J. Funct. Anal.* **75** (1987) 161–187
- [Pe3] Perry P, The Laplace operator on a hyperbolic manifold, II. Eisenstein series and the scattering matrix, *J. Reine Angew. Math.* **39** (1989) 67–91
- [Pe4] Perry P, The Selberg zeta function and a local trace formula for Kleinian groups, *J. Reine Angew. Math.* **40** (1990) 116–152
- [Pe5] Perry P, The Selberg zeta function and scattering poles for Kleinian groups, *Bull. Am. Math. Soc. (N.S.)* **24** (1991) 327–333
- [PSS] Perry P, Sigal I M and Simon B, Spectral analysis of  $N$ -body Schrödinger operators, *Ann. Math.* **114** (1981) 519–567
- [PS1] Phillips R S and Sarnak P, On cusps forms for co-finite subgroups of  $PSL(2, \mathbb{R})$ , *Invent. Math.* **80** (1985) 339–364
- [PS2] Phillips R S and Sarnak P, Perturbation theory for the Laplacian on automorphic functions, *J. Am. Math. Soc.* **5** (1992) 1–32
- [RS1] Reed M, Simon B, *Methods of modern mathematical physics, Functional Analysis* (1980) (New York: Academic Press) Vol. I
- [RS2] Reed M and Simon B, *Methods of modern mathematical physics, Fourier analysis and self-adjointness* (1981) (New York: Academic Press) Vol. II
- [RS4] Reed M and Simon B, *Methods of modern mathematical physics, Analysis of operators* (1978) (New York: Academic Press) Vol. IV
- [S] Sullivan D, The density at infinity of a discrete group of hyperbolic motions, *I.H.E.S. Publ. Math.* **50** (1979) 171–202
- [Se] Selberg A, Göttingen Lectures, 1954 (for a published proof, see [LP5]).

- [SWY] Schoen R, Wolpert S and Yau S T, Geometric bounds on the low Eigenvalues of a compact surface, *Proc. Sympos. Pure Math* (1980) (Providence, R.I.: American Mathematical Society) **36** 279–285
- [Ta] Taylor M E, *Pseudodifferential operators* (1981) (Princeton, N.J.: Princeton University Press)
- [T] Terras A, *Harmonic analysis on symmetric spaces and applications*, I. (1985) (New York: Springer-Verlag)
- [Th] Thurston W, *The topology and geometry of three-manifolds* Princeton, N.J.: Princeton University Lecture Notes, 1977
- [Ti] Titchmarsh E C, *Eigenfunction expansions, Part 1* (1962) (Oxford: Clarendon Press)
- [Z] Zworski M, Counting scattering poles, in *Spectral and Scattering Theory*, (ed.) M Ikawa, 1994 (Marcel Dekker Publishing Co.)

# Inverse spectral theory for Jacobi matrices and their almost periodicity

ANAND J ANTONY and M KRISHNA\*

School of Mathematics, SPIC Science Foundation, Madras 600017, India

\*Institute of Mathematical Sciences, Taramani, Madras 600 113, India

**Abstract.** In this paper we consider the inverse problem for bounded Jacobi matrices with nonempty absolutely continuous spectrum and as an application show the almost periodicity of some random Jacobi matrices. We do the inversion in two different ways. In the general case we use a direct method of reconstructing the Green functions. In the special case where we show the almost periodicity, we use an alternative method using the trace formula for points in the orbit of the matrices under translations. This method of reconstruction involves analyzing the Abel-Jacobi map and solving of the Jacobi inversion problem associated with an infinite genus Riemann surface constructed from the spectrum.

**Keywords.** Jacobi matrices; inverse theory; almost periodicity.

## 1. Introduction

In this paper we address the question of recovering a Jacobi matrix

$$Hu(n) = a_n u(n+1) + b_n u(n) + a_{n-1} u(n-1), \quad u \in l^2(\mathbb{Z}) \quad (1)$$

with  $a_n \geq 0$  and  $b_n$  real, from its spectrum  $\Sigma$ , assuming it to be a compact set, and also the question of the almost periodicity of the sequences  $a_n$  and  $b_n$  constructed from the spectrum. We consider the case, when the real part of the boundary values of the Green function for the vector  $\delta_0$  vanish almost everywhere on the spectrum. In this case we show that there is no uniqueness even when the Dirichlet eigenvalues of the half-line problems are specified.

We use this theory to prove the almost periodicity of some random Jacobi matrices with the spectrum having a band structure. The motivation for this work comes from the work on periodic Jacobi matrices and the inverse theory for Schrödinger operators. There is extensive work on inverse spectral theory in the sense of recovering the operators from given spectral quantities, for periodic Schrödinger operators in the literature, the work of McKean–Moerbeke [31], McKean–Trubowitz [32], Trubowitz [38] being the some of these. There is also the work of Dubrovin–Matveev–Novikov [9], Levitan [27], [28], [29] for almost periodic potentials. In a general framework of ergodic potentials, the inverse spectral theory was initiated by Kotani, who showed the existence of ergodic Schrödinger operators associated with a class of spectral functions, and also discussed classical integrable systems in this framework in a series of papers ([18]–[21]). These inversion results of Kotani produced classes of potentials, while the pointwise information, and the nature of the isospectral class obtained were discussed in Kotani–Krishna [23] for ergodic potentials and Craig [8] for a very general class of reflectionless potentials. In the

discrete case there are several analogues of the above, for periodic examples, in the works of Kac–Moerbecke [14], [15], Dubrovin–Matveev–Novikov [9], Toda [37]. In the case of random Jacobi matrices, Carmona–Kotani [5] obtain an invariant probability measure associated with some spectral functions. With these examples in mind we wanted to address two questions, one is to invert a Jacobi matrix from its spectrum and the other is to identify the situations when the resulting matrix is almost periodic. For the case of finite band spectra we [2] proved almost periodicity of the Jacobi matrices.

More recently there has been some interest in the inverse spectral theory, and a very general set-up is proposed by Gesztesy–Simon, see the announcement [12], who identified the xi function as the central object for inverse theories. They also give a new proof for existence of absolutely continuous spectrum for almost mathieu operators with small coupling. There is also the work of Knill [17] on isospectral deformation for random Jacobi matrices. As this paper was nearing completion we received the beautiful lecture notes of Simon [36] on the applications of rank one perturbations to inverse theory, with some constructions similar to what we do in §2.

We present below the assumptions we make (needed for §2) on the spectrum and on the random potential.

*Assumption 1.*  $\Sigma$  is a compact subset of  $\mathbb{R}$  of positive Lebesgue measure.

We can write  $\Sigma$  as the complement of the disjoint union of a countable number of open intervals, which we call the gaps and fix notation as follows.

$$\Sigma = \mathbb{R} \setminus [I(-\infty) \cup I(+\infty) \cup \bigcup_{i=1}^{\infty} I_i], \quad (2)$$

where, we take  $\tau_0 = \inf \Sigma$  and  $\tau_{\infty} = \sup \Sigma$ ,

$$I(-\infty) = (-\infty, \tau_0), \quad I(\infty) = (\tau_{\infty}, \infty), \quad I_i = (\tau_{2i-1}, \tau_{2i}).$$

We denote  $I_k < I_i$  to mean  $I_k$  is to the left of  $I_i$ . This gives an ordering of the gaps which is convenient for deducing the properties of some of the functions to be introduced in §3 which will be used later. For a given gap  $I_i$  we denote the infimum of its distance from other gaps as  $4s_i$ , i.e.  $4s_i = \inf_{k \neq i} \text{dist}(I_i, I_k)$ . We call  $s_0, s_{\infty}$  the distances of  $I(-\infty)$  and  $I(\infty)$  from the nearest gaps. We also set  $q_i = l_i/s_i$  and we take without loss of generality that  $q_i < 1$ . We also note that assumption (1) implies that  $l_i$  and  $s_i$  are summable. With these notations the following assumptions on the spectrum are used from §3 onwards.

*Assumption 2.* Let  $\Sigma$  satisfy assumption (1) and in addition suppose,

1.  $s_i > 0$  for all  $i = 1, 2, \dots$  and for  $i = 0, \infty$ .
2.  $\sum_{i=1}^{\infty} q_i < \infty$  and  $\sum_{i=1}^{\infty} q_i/s_i < \infty$ .
3. Let  $\mathcal{E}$  denote the set of accumulation points of the boundary points of  $\Sigma$  and  $\mathcal{E}_1$  the set of accumulation points of  $\mathcal{E}$ . Then we assume that  $\mathcal{E}_1$  is a finite set.

For the results of the final section we assume that the probability measure  $\mathbb{P}$  (see §5 for definitions) satisfies,



With these assumptions and notations our main theorem is the following.

**Theorem 1.1.** *Let  $\mathbb{P}$  satisfy assumption (3) and let the spectrum  $\Sigma$  satisfy the assumptions (1, 2). Then every point  $\omega$  in the support of  $\mathbb{P}$  is an almost periodic sequence, in the sense that  $b_n^\omega$  and  $a_n^\omega$  are almost periodic sequences for each  $\omega$ .*

### 1.1 Ideas, strategies and limitations

In this subsection we discuss the questions we addressed in this paper and provide some clarifications of the assumptions we make use of and the results we obtain using our method and also discuss some of the limitations of our proofs.

In the Schrödinger case (see Kotani–Krishna [23] or Craig [8]) the Dubrovin equation, namely,  $d\xi_i(t)/dt = W_i(\xi(t))$  for the zeros  $\xi_i(t)$  of the Green functions  $g_\lambda(t, t)$  in the gaps  $I_i$ , came for free, from the differential equation

$$2(g_\lambda''(x, x) - 2(q - \lambda)g_\lambda(x, x))g_\lambda(x, x) - g_\lambda'^2 + 1 = 0$$

satisfied by the Green functions in the resolvent set. This equation was used by Moser [30], McKean–Moerbeke [31] and McKean–Trubowitz [32] in the context of inverse spectral theories. To solve for  $\{\xi_i(t), t \in \mathbb{R}\}$ , it is enough to get good estimates for  $W_i(\xi)$  in the gaps (allowing for example Cantor like spectra) as done by Craig [8]. In the Jacobi case information in the gaps alone is not sufficient, we need to know even the behaviour of some spectral functions in the bands also, even in the case of reflectionless potentials, where all the Green functions  $g_\lambda(n, n)$  have vanishing real parts in the spectrum.

In our earlier work [2] on random Jacobi matrices with finite band absolutely continuous spectrum, we showed that the zeros  $\xi_i(n)$ , in the gaps, of the Green functions  $g_\lambda(n, n)$  are related to the theta function of a finite genus hyperelliptic compact Riemann surface via the relation,

$$\sum_{i=1}^N \xi_i(n) = \sum_{i=1}^N C_{iN} D_i \ln \left\{ \frac{\Theta(nc + d)}{\Theta((n+1)c + d)} \right\},$$

where  $c$  and  $d$  are real  $N$  dimensional vectors and  $D_i$  is the derivative in the  $i$ th direction. The theta function is defined via the period matrix,  $i\tau$ ,  $\tau$  being positive definite  $N$  dimensional matrix, by

$$\Theta(z) = \sum_{m \in \mathbb{Z}^N} \exp(2\pi i \langle m, z \rangle) \exp(-\langle m, \tau m \rangle)$$

so that the almost periodicity of the theta function with real argument can be used to conclude the almost periodicity of the diagonal entries of the Jacobi matrix, through the trace formula. In the current case however the Riemann surface is of infinite genus and it is simpler to use the method of Levitan for showing the almost periodicity directly which is as follows.. Note that the linear flow  $tc + d$  projected to  $\mathbb{R}^K/\mathbb{Z}^K$  is almost periodic, for fixed vectors  $c$  and  $d$  in  $\mathbb{R}^K$ . If we consider a Lipschitz continuous  $\mathbb{Z}^K$  periodic map  $f$  of  $\mathbb{R}^K$  to itself, then we show that each of the co-ordinates of the image  $x(t) = f(tc + d)$ , is an almost periodic function. Hence the sum of the co-ordinates  $\Sigma x_i(t)$  is also an almost periodic function. In the infinite dimensional

setting one needs to truncate and argue, essentially using the finite dimensional result. One of the harder parts of the proof requires showing that the zeros  $\xi_i(n)$  of the Green functions  $g_\lambda(n, n)$  in the gaps  $I_i$ , which satisfy an analogue of the Dubrovin equation of the continuous case, are indeed of the above form.

In the section on Inverse spectral theory, we show how to reconstruct a Jacobi matrix from the spectrum when the Green function for the vector  $\delta_0$  of the Jacobi matrix has vanishing real part almost everywhere on the spectrum, which we assume to be a compact subset of the reals of positive Lebesgue measure. The method of this section obtains a Jacobi matrix given the spectrum along with a collection of points one each in the gaps, which will serve as the Dirichlet eigen values for a half-line problem. However the reconstruction does not give a unique answer. The Green function  $g_\lambda(0, 0)$  of the Jacobi matrix and those of the half-line Dirichlet problems  $m^\pm(\lambda)$  are related by  $g_\lambda(0, 0) = -1/(m^+(\lambda) + m^-(\lambda) + \lambda - b_0)$ . This equation is used to conclude that the zero of the Green function in the gaps will be eigenvalues of one of the half-line problems. Hence there are several Jacobi matrices with the same spectrum and the zeros of the Green function in the gaps. In the Schrödinger case the source of non-uniqueness was just this for a general class of reflectionless potentials. However in the discrete case there are two additional sources of non-uniqueness even when the Dirichlet eigen values of the half-line problems are specified. One is that even if the Jacobi matrix  $H$  has purely absolutely continuous spectrum  $\Sigma$ , the half-line operators  $H^\pm$  may have some singular spectrum in  $\Sigma$ . Alternately, even if  $H^\pm$  have no singular spectrum in  $\Sigma$ , the spectral measures of  $H^\pm$  associated with the vectors  $\delta_{\pm 1}$  restricted to  $\Sigma$  may be different. In the last section we will see that for an ergodic Jacobi matrix with purely absolutely continuous spectrum given by a band structure both these sources of non-uniqueness will be absent.

Starting with §3, we concentrate on a special class of spectra and set up the machinery required for proving almost periodicity of some Jacobi matrices. In §3, we discuss an interpolation theorem for a class of analytic functions associated with the spectrum of a Jacobi matrix.

In the subsequent section, we discuss the Riemann surface and the Abel–Jacobi map on a class of divisors. We compute the image of the divisors in this class and for subsequent applications, extend the map to an infinite dimensional Banach space. The proposition (4.3) reminds one of the classical theorem of Abel, for compact surfaces, that the principal divisors of degree zero are precisely those whose image in the Jacobi variety under the Abel–Jacobi map is 0. It is too tempting from our proofs to conclude a similar theorem for the surface  $\mathcal{R}$ . But unfortunately it is not entirely clear even when the set  $\mathcal{E}$  is finite. The reason is that if we consider an arbitrary meromorphic function, on the Riemann surface  $\mathcal{R}$  its behaviour on the bounding curves of the approximants is unclear, even under the simplifying assumptions used on the spectrum. Our proof of the computation of the image of the Abel–Jacobi map is essentially proving the bilinear relations of Riemann (though we do not state the relations in this paper) for a class of meromorphic differentials and these relations are valid only in a very special sense, and hence only for a special class of meromorphic differentials, in our context.

We would like to add a few more words about the Abel–Jacobi map whose properties might appear mysterious if not clarified. The set up and the ideas we use are from the beautiful papers of Levitan [27, 28, 29] who was working with

Schrödinger operators. We believe that almost all the points we describe below are essentially contained in his work though they might not have been stated explicitly.

Classically (that is in the case of a compact Riemann surface) it is a map from the divisors on the surface to the Jacobi variety (a compact object). Even in our case the Jacobi variety is compact and we consider only some divisors in "real position" (see McKean-Trubowitz [32] for more detail on these) and hence the real part of the Jacobi variety. However the Jacobi variety is not a linear space. It is convenient to work on a linear space to be able to use the inverse function theorem, for Jacobi inversion, in this infinite dimensional setting.

The Abel-Jacobi map is set up on an infinite dimensional real Banach space (with its norm chosen using the geometrical conditions on the spectrum). In this framework, the Abel-Jacobi map may be **non-differentiable**. (The problem is that the family of functions  $f_{i,s}$ , introduced in the equation (45) is not necessarily equicontinuous family, so that the differentiability of  $\mathcal{A}_1$  is not at all clear). It is however Lipschitz continuous and its inverse is also Lipschitz continuous and this is enough for the proofs of almost periodicity of some spectral parameters. To achieve this we split the Abel-Jacobi map into its "diagonal" and "off-diagonal" parts. The "diagonal" part is a Lipschitz continuous bijection and the "off-diagonal" part is a differentiable, periodic, (with its periods coming from the lattice of  $(\pi)$  integral points of the Banach space) and a compact map on the Banach space. This is the reason for introducing the auxiliary map  $\mathcal{B}$ .

We use throughout this paper the notation  $\mathbb{Z}^\pm$  for  $\{\pm 1, \pm 2, \pm 3, \dots\}$ , unless stated otherwise.

## 2. Inverse spectral theory

In this section we start with a compact subset  $\Sigma$  of  $\mathbb{R}$  of positive Lebesgue measure and obtain a Jacobi matrix which has  $\Sigma$  as the essential closure of its absolutely continuous spectrum. The procedure involves constructing a Herglotz function which is a likely candidate for the Green function  $g_\lambda(0, 0)$ , by specifying its argument on the set  $\Sigma$  together with a given asymptotic behaviour. We then look at the inverse of the function so constructed, determine its poles outside the set  $\Sigma$  and construct two Herglotz functions which will be the Green functions of Jacobi matrices on square integrable sequences on the left and right half-lines. These will have simple spectrum and from these the full Jacobi matrix is recovered.

There is a long history for the inverse problem for a Jacobi matrix. There are the classical works of Akhiezer [1], Kac-Moerbecke [14], [15], Moerbeke [39] Dubrovin-Matveev-Novikov [9] done for the half-line problem and for periodic sequences. The latest such inversion on the half-line is in the work of Rajaram Bhat-Parthasarathy [34], in a different context. Some of the ideas used here in the construction of the Herglotz functions associated with a given set can be found in the works of Kotani [18] [21], Kotani-Krishna [23], Craig [8], Levitan [27] [28], [29], Gesztesy-Holden-Simon-Zhao [12] and Gesztesy-Simon [13]. In addition, the book of

function to be constructed. Henceforth we use the following notation,

$$\begin{aligned}\Pi &= \mathbb{C} \setminus \Sigma, \quad \Psi = \prod_{i=1}^{\infty} \bar{I}_i \\ S_{\xi} &= \{\xi_1, \xi_2, \dots, \xi_n, \dots\}, \quad \xi \in \Psi \\ \mathcal{O} &= \bigcup_{i=1}^{\infty} I_i, \quad \mathcal{O}_{\xi} = \mathcal{O} \cap S_{\xi}.\end{aligned}$$

where  $\xi$  is a sequence while  $S_{\xi}$  is a subset of the complex plane and the line above the set  $I_i$  indicates the closure of the set. For  $\xi$  in  $\Psi$ , consider a partition  $S_{\xi}^{+} \cup S_{\xi}^{-}$  of  $S_{\xi}$  into disjoint subsets  $S_{\xi}^{\pm}$ , we write

$$\mathcal{O}_{\xi}^{\pm} = \mathcal{O} \cap S_{\xi}^{\pm}. \quad (3)$$

Given this information we can proceed to construct a class of functions as follows. In the following for simplicity of notation we set

$$k(\lambda, x) = \frac{1}{(x - \lambda)} \quad \text{and} \quad K(\lambda, x) = k(\lambda, x) - \frac{x}{1 + x^2} \quad (4)$$

and choose the square root occurring in the log in the following lemma so that it is positive on  $(-\infty, 0)$ .

*Lemma 2.1. Consider  $\Sigma$  as in assumption 1 and let  $\xi \in \Psi$ . Then there is a Herglotz function  $F_{\xi}(\lambda)$  such that*

$$\begin{aligned}F_{\xi}(\lambda) &= \int k(\lambda, x) d\sigma(x) + \frac{1}{2} \log \frac{1}{(\tau_0 - \lambda)(\tau_{\infty} - \lambda)}, \quad \lambda \in \mathbb{C}^{+} \\ \int k(\lambda, x) d\sigma(x) &= \frac{1}{2} \sum_{i=1}^{\infty} \log \frac{(\xi_i - \lambda)}{(\tau_{2i} - \lambda)} + \log \frac{(\xi_i - \lambda)}{(\tau_{2i-1} - \lambda)}\end{aligned} \quad (5)$$

with  $\sigma$  a signed absolutely continuous measure of finite total variation. Further  $F_{\xi}(\lambda) \rightarrow \log(-1/\lambda)$  as  $\lambda \rightarrow \infty$  and

$$\operatorname{Im} F_{\xi}(x + i0) = \begin{cases} \frac{\pi}{2} & \text{a.e. } x \in \Sigma \\ \pi & \text{on } (-\infty, \xi_i) \cap I_i \quad \forall i = 1, 2, \dots, \text{ and on } I(\infty) \\ 0 & \text{on } (\xi_i, \infty) \cap I_i \quad \forall i = 1, 2, \dots, \text{ and on } I(-\infty). \end{cases}$$

The sum in (5) converges compact uniformly in  $\Pi$ .

*Proof.* Consider a non-negative bounded function, specified almost everywhere by,

$$\xi(x) = \begin{cases} \frac{\pi}{2} & \text{a.e. } x \in \Sigma \\ \pi & \text{on } (-\infty, \xi_i) \cap I_i \quad \forall i = 1, 2, \dots, \text{ and on } I(\infty) \\ 0 & \text{on } (\xi_i, \infty) \cap I_i \quad \forall i = 1, 2, \dots, \text{ and on } I(-\infty). \end{cases}$$

The  $\int \xi(x) dx / (1 + x^2) < \infty$ . Therefore for each  $c \in \mathbb{R}$ , the function

$$F_\xi(\lambda) = c + \frac{1}{\pi} \int K(\lambda, x) \xi(x) dx \quad (6)$$

is Herglotz. It is convenient sometimes to write  $F$ , as done by Craig [8], in terms of a signed measure. Therefore we choose another function,

$$\phi(x) = \frac{\pi}{2} \quad \forall x \in [\tau_0, \tau_\infty] \quad \text{and} \quad \pi \quad \forall x \in (\tau_\infty, \infty) \quad (7)$$

taken to be 0 otherwise, and define

$$G_\xi(\lambda) = \frac{1}{\pi} \int K(\lambda, x) \phi(x) dx. \quad (8)$$

We find that  $G$  is also Herglotz, since  $\int \phi(x) dx / (1 + x^2) < \infty$ . We then write

$$F_\xi = F_\xi - G_\xi + G_\xi$$

and choose the number  $c$  so that  $F_\xi$  has the representation stated in the Lemma. The asymptotic behaviour for  $F_\xi$  comes from the second term in the expression for  $F_\xi$  given in the statement of the lemma, since the measure  $\sigma$  is finite. For clarity we write the measure  $\sigma$  below.

$$d\sigma(x) = \sum_{i=1}^{\infty} \frac{1}{2} (\chi_{(\tau_{2i-1}, \xi_i)}(x) - \chi_{(\xi_i, \tau_{2i})}(x)) dx$$

We get the series expansion for  $F_\xi$  by integration by parts. Let  $S$  be a compact subset of  $\Pi$ . Then there is an  $\varepsilon > 0$ , such that  $\text{dist}(S, \Sigma) > \varepsilon$ . We have the following estimates, for any  $i = 1, 2, \dots$  and  $\xi_i \neq \tau_{2i-1}$  or  $\tau_{2i}$ . Clearly the estimate is trivial for  $\xi_i = \tau_{2i-1}$  or  $\tau_{2i}$ .

$$\left| \frac{(\xi_i - \lambda)}{(\tau_{2i-1} - \lambda)} \right| \leq 1 + \left| \frac{(\xi_i - \tau_{2i-1})}{(\tau_{2i-1} - \lambda)} \right| \leq 1 + \left| \frac{l_i}{\varepsilon} \right| \quad (9)$$

and

$$\left| \frac{(\xi_i - \lambda)}{(\tau_{2i} - \lambda)} \right| \leq 1 + \left| \frac{(\xi_i - \tau_{2i})}{(\tau_{2i} - \lambda)} \right| \leq 1 + \left| \frac{l_i}{\varepsilon} \right|. \quad (10)$$

where  $l_i$ 's are the gap lengths  $|I_i|$ . Their sum converges since, all  $I_i$ 's are a disjoint union of intervals contained in  $[\tau_0, \tau_\infty]$ . This shows the convergence, uniformly on compacts of  $\Pi$ , of the sum in equation (5).  $\square$

By exponentiating  $F_\xi$  of the above Lemma, we get the following proposition, where we define the formal products

$$P_\xi(\lambda) = \prod (\lambda - \xi_i) \quad \text{and} \quad R(\lambda) = (\lambda - \tau_0)(\lambda - \tau_\infty) \prod_i (\lambda - \tau_i).$$

appendix, that the measure  $\mu$  in the following representation has an absolutely continuous component supported on  $\Sigma$ .

## PROPOSITION 2.2

Consider  $\Sigma$  as in Assumption (1) and let  $\xi \in \Psi$ . Then there is a unique Herglotz function  $h_\xi$  analytic in  $\Pi$  satisfying the following properties.

1.  $h_\xi$  has simple zeros on  $\mathcal{O}_\xi$ .
2.  $h_\xi = \frac{-1}{\lambda} + O\left(\frac{1}{\lambda^2}\right)$  as  $\lambda \rightarrow \infty$ .
3.  $h_\xi(x + i0)$  has the following values on  $\mathbb{R}$ .

$$\operatorname{Re} h_\xi(x + i0) = 0 \quad \text{a.e. on } \Sigma, \quad \operatorname{Im} h_\xi(x + i0) = 0, \quad \forall x \in \mathbb{R} \setminus \Sigma$$

$$\operatorname{Re} h_\xi(x + i0) > 0, \quad x \in I(-\infty) \cup_{i=1}^\infty (\xi_i, \infty) \cap I_i$$

$$\operatorname{Re} h_\xi(x + i0) < 0, \quad x \in I(\infty) \cup_{i=1}^\infty (-\infty, \xi_i) \cap I_i$$

$h_\xi$  also has the following Herglotz representation,

$$h_\xi(\lambda) = \int k(\lambda, x) d\mu(x) \quad (11)$$

with  $\mu$  supported on  $\Sigma$  with the absolutely continuous component of  $\mu$  having essential support  $\Sigma$ . The product representation

$$h_\xi(\lambda) = \frac{P_\xi(\lambda)}{\sqrt{R(\lambda)}} \quad (12)$$

is also valid with the products converging uniformly on compacts of  $\Pi$ .

*Proof.* We consider the function

$$h_\xi(\lambda) = \exp(F_\xi(\lambda)),$$

where  $F_\xi$  is as in lemma (2.1). Then the stated properties of  $h_\xi$  follow from those of  $F_\xi$ , the equation (5) and the estimates (9) and (10) used in concluding the convergence of the sum representing  $F_\xi$ .  $\square$

The following lemma will be crucial for solving the inverse problem. The lemma gives a Herglotz representation for the inverse of  $h_\xi$  given above.

**Lemma 2.3.** *The function  $g_\xi = -h_\xi^{-1}$  is Herglotz and has the representation*

$$g_\xi(\lambda) = \lambda + d_\xi + \int k(\lambda, x) dv(x) \quad (13)$$

where  $v$  is a finite positive measure with support  $\Sigma \cup \mathcal{O}_\xi$  with the absolutely continuous part having support in  $\Sigma$ . Its singular part in  $\mathcal{O}_\xi$  is pure point.

*Proof.* It is clear that  $g_\xi$  defined above is Herglotz, since  $h_\xi$  is Herglotz and has the

stated representation, from proposition (A.2) (4), since the support of the representing measure is compact and from the definition of  $g_\xi$  its behaviour at infinity is like  $\lambda^{-1}$ . We let  $\nu$  be the finite positive measure given by the Herglotz representation theorem. We note that equation (12) gives the product representation

$$g_\xi = - \frac{\sqrt{R(\lambda)}}{P_\xi(\lambda)} \quad (14)$$

with the product converging uniformly on compacts of  $\Pi$ . It is also clear that from the estimates of (5),  $g_\xi$  has a simple pole in  $I_i$ , whenever  $\xi_i \in I_i$ . As for the set  $\Sigma$ , the limits  $\text{Im } g_\xi(x + i0)$  exist finitely and are positive, at every point where  $h_\xi(x + i0)$  has a positive finite imaginary part. Since this happens a.e. on  $\Sigma$ , the same is true for  $g_\xi$ . This shows that the absolutely continuous part of  $\nu$  is non zero and has essential support  $\Sigma$ .  $\square$

We write  $\nu$  on  $\Sigma$  and  $\mathcal{O}_\xi$  as  $\nu_1$  and  $\nu_2$  respectively, consider the partition of (3) and define  $\nu_2^\pm \equiv \nu_2|_{\mathcal{O}_\xi^\pm}$ . Given these we consider some positive measures,  $\nu_{1\pm}$  so that  $\nu_1 + \nu_{1-} = \nu_1$  as measures and define

$$\nu^\pm = \nu_{1\pm} + \nu_2^\pm \quad \text{and} \quad c^\pm = \int d\nu^\pm. \quad (15)$$

We then have the following theorem, where we denote by  $(\cdot)^{-\text{ess}}$  the closure up to sets of Lebesgue measure zero.

**Theorem 2.4.** Consider the set  $\Sigma \cup \mathcal{O}_\xi$ ,  $\Sigma$  as in assumption (1) and the measures  $\nu^\pm$ . Then there exist unique Jacobi matrices  $H^\pm$  on  $l^2(\mathbb{Z}^+)$  with simple spectra such that

$$c^\pm (H^\pm - \lambda)^{-1}(\pm 1, \pm 1) = \int k(\lambda, x) d\nu^\pm(x).$$

In particular the absolutely continuous spectrum of  $H^\pm$  is  $\Sigma^{-\text{ess}}$  and they have eigenvalues on  $\mathcal{O}_\xi^\pm$  respectively.

*Proof.* We shall consider the  $+$  case the other one is similar. Consider  $L^2\left(\Sigma \cup \mathcal{O}_\xi, \frac{\nu^+}{c^+}\right)$

and let  $M$  be the operator of multiplication by  $x$  on this space. Thus  $M$  is a bounded self-adjoint operator. The constant function 1 on  $\Sigma \cup \mathcal{O}_\xi^+$  together with the monomials  $x^m$ ,  $m = 1, 2, \dots$  form a basis for the above Hilbert space, therefore by the Gram-Schmidt procedure we can get an orthonormal basis  $\{e_n\}$ ,  $n = 1, 2, \dots$ , out of these. We write the matrix elements of  $M$  as  $M_{ij} = \langle e_i, Me_j \rangle$  in this basis and see that for  $i > j + 1$ ,  $M_{ij}$  is zero, since  $Me_j$  is at most a polynomial of degree  $j + 1$ . On the other hand since  $\langle e_i, Me_j \rangle = \overline{\langle e_j, Me_i \rangle}$ , by the self-adjointness of  $M$ , it follows that for  $i < j - 1$  also  $M_{ij}$  is zero showing that  $M$  is tridiagonal in this basis. Clearly by the self-adjointness of  $M$  the diagonal entries are real and the off diagonal entries can be chosen to be positive, by choosing the phases of the vectors  $e_n$  appropriately.

We take the unitary isomorphism  $U$  from  $L^2\left(\Sigma \cup \mathcal{O}_\xi, \frac{\nu^+}{c^+}\right)$  to  $l^2(\mathbb{Z}^+)$  taking the

basis  $\{e_n\}$  to the canonical basis  $\delta_n$ . We define  $H^+ = U M U^{-1}$ . Then  $\langle e_n, M e_m \rangle = \langle \delta_n, H^+ \delta_m \rangle$ . This gives us the uniqueness. We write the matrix elements of  $H^+$  as

$$H_{i,i+1}^+ = H_{i+1,i}^+ = a_i \quad \text{and} \quad H_{ii}^+ = b_i \quad i \in \mathbb{Z}^+.$$

We construct  $H^-$  on  $l^2(\mathbb{Z}^-)$  similarly using  $v^-$ . With these definitions the operators  $H^\pm$  act on  $l^2(\mathbb{Z}^\pm)$  as

$$(H^\pm)u(n) = a_n u(n+1) + b_n u(n) + a_{n-1} u(n-1), \quad |n| \neq 1$$

and

$$(H^+)u(1) = a_1 u(2) + b_1 u(1), \quad \text{and} \quad (H^-)u(-1) = b_{-1} u(-1) + a_{-2} u(-2).$$

which proves the lemma.  $\square$

We also have by construction that,  $M^\pm$  defined by

$$M^\pm(\lambda) \equiv \int k(\lambda, x) dv^\pm \quad \text{satisfies} \quad M^\pm(\lambda) = c^\pm (H^\pm - \lambda)^{-1} (\pm 1, \pm 1) \quad (16)$$

for  $\lambda \in \mathbb{C}^+$ . Clearly the above expression can also be rewritten in the special case when  $v_{1\pm} = \frac{1}{2} v_1$ , as

$$M^\pm(\lambda) = \frac{1}{2} \int k(\lambda, x) dv \pm \int k(\lambda, x) d(v_2^+ - v_2^-). \quad (17)$$

For reconstructing the Jacobi matrix we set, using the function  $g_\xi$  of lemma (2.3) and positive square roots of  $c^\pm$ ,

$$a_0 = \sqrt{c^+}, \quad a_{-1} = \sqrt{c^-}, \quad \text{and} \quad b_0 = d_\xi.$$

and note that  $b_0$  is real since  $g_\xi$  is Herglotz. We prefer to state the theorem by fixing the set of Dirichlet eigen values for the half-line problems to emphasize the non-uniqueness present in the problem.

**Theorem 2.5.** Consider the set  $\Sigma$  as in assumption (1) and the sets  $\mathcal{O}_\xi^\pm$  associated with a point  $\xi \in \Psi$  and a partition  $S_\xi^\pm$ . Then there exists a Jacobi matrix  $H$  with

$$(H - \lambda)^{-1}(0, 0) = h_\xi(\lambda)$$

The absolutely continuous spectrum of  $H$  is  $\Sigma^{-\text{ess}}$ , has multiplicity 2 and the half-line operators  $H^\pm$  have eigen values on  $\mathcal{O}_\xi^\pm$ . Such a  $H$  is unique if the spectral measures  $v^\pm$  of  $H^\pm$  for the vectors  $\delta_{\pm 1}$  are also specified in which case we have

$$c^\pm (H^\pm - \lambda)^{-1} = \int k(\lambda, x) dv^\pm$$

*Proof.* Given  $\Sigma$  we construct  $h_\xi$  as in Proposition (2.2),  $g_\xi$  as in Proposition (2.3) and consider  $v_1$ . Decompose  $v_1$  into its absolutely continuous  $v_{1ac}$  and its singular  $v_{1s}$  parts. Consider any partition  $\rho_1 + \rho_2$  of unity and consider, the measures

$$v_{1+} = \rho_1 v_{1ac}, \quad v_{1-} = \rho_2 v_{1ac} + v_{1s}$$



constructed in theorem (2.4) and define the operator  $H$  on  $l^2(\mathbb{Z})$  using  $v^\pm$  by

$$(Hu)(n) = a_n u(n+1) + b_n u(n) + a_{n-1} u(n-1).$$

It is clearly a bounded self-adjoint operator and has  $\Sigma^{-\text{ess}}$  as its absolutely continuous spectrum of multiplicity 2, since  $H$  and  $H^+ \oplus H^-$  differ by a (finite rank and hence) trace class operator and  $H^+ \oplus H^-$  has  $\Sigma^{-\text{ess}}$  as its absolutely continuous spectrum with multiplicity 2. The remaining statement is clear by construction. Its uniqueness follows from the unique reconstruction of the operators  $H^\pm$  from  $v^\pm$  together with the determination of the numbers  $a_0, a_{-1}$  and  $b_0$  are uniquely obtained from  $v^\pm$  and  $h_\xi$ . The other properties are clear as in the previous proposition.  $\square$

*Remark 2.6.* The above theorem is a reformulation of the traditional way of stating the inverse theorem in terms of associating the operators  $H^\pm$  to spectral parameters  $\{\tau_i, \xi_i, \sigma_i\}$ , with  $\sigma_i$  given values  $\pm 1$  whenever  $\xi_i$  is the eigen value of  $H^\pm$  in the gaps. The formulation we give here seems better. We should note here that there is a great deal of non-uniqueness in the above construction, even when the measures  $v_{2\pm}$  are fixed. The special case  $v_{1+} = v_{1-} = \frac{1}{2}v_1$  is valid for a class of ergodic Jacobi matrices, as we shall see in the last section. The asymmetry  $a_0 = \int v^+$  while  $a_{-1} = \int v^-$  comes from the definition of  $m$ -functions of the Jacobi matrix, see Simon [35] for example.

### 3. Interpolation theorem

In this section we consider a class of analytic and meromorphic functions on the complement of the spectrum  $\Sigma$  in the complex plane that would be candidates for the Green functions of Jacobi matrices. We impose further restrictions on the set  $\Sigma$  considered in the last section, and take the other quantities associated with  $\Sigma$  as before. In this setting we consider a class of analytic functions  $\mathcal{H}_{\infty-1}$  which have zeros one each in all the gaps except one and prove an interpolation theorem that lets us recover any of the functions in this class from their values at a point each in each of the gaps. This class of functions will give the differentials of the first kind on a Riemann surface to be constructed later. We also construct a family of functions from  $\mathcal{H}_{\infty-1}$  with the property that their integral over each of the gaps except one vanishes and in the exceptional gap it is normalized. This family of functions will give normalized differentials of the first kind on the Riemann surface of the next section.

We start with some elementary estimates in the following lemma.

*Lemma 3.1.* Suppose  $\Sigma$  satisfies assumptions (2). Then the following bounds are valid for any  $\lambda$ . We set  $\text{dist}(\lambda, I_i)$  to be  $d_i$ ,  $i = 1, 2, \dots$ ,  $\text{dist}(\lambda, \tau_0) = d_0$  and  $\text{dist}(\lambda, \tau_x) = d_x$ .

$$\left| \frac{(\xi_i - \lambda)}{(\tau_{2i-1} - \lambda)} \right| \leq 1 + \left| \frac{(\xi_i - \tau_{2i-1})}{(\tau_{2i-1} - \lambda)} \right| \leq 1 + \left| \frac{l_i}{d_i} \right| \quad (18)$$

and

$$\left| \frac{(\xi_i - \lambda)}{(\tau_{2i} - \lambda)} \right| \leq 1 + \left| \frac{(\xi_i - \tau_{2i})}{(\tau_{2i} - \lambda)} \right| \leq 1 + \left| \frac{l_i}{d_i} \right|. \quad (19)$$

$$\left| \frac{1}{\tau_{2i-1} - \lambda} \right| \leq \left| \frac{1}{\tau_{2i} - \lambda} \right| \quad (20)$$

*Proof.* The proof is trivial once we allow both side to be infinite.  $\square$

We would like to recover a class of functions from their values at a given set of points as in the case of a polynomial of degree  $n$  which can be recovered from its values on a set of  $n + 1$  points. This is the main idea of the interpolation theorem. Towards stating the interpolation, we consider a collection of points one each from the closure of each gap, or the set of points coming from all gaps except one. The notation  $\infty - 1$  is strange, but as in the work of McKean-Trubowitz [32], it is natural. Associated with these points we define classes of analytic functions below. To start with we define a set, where the union is the disjoint union, given by

$$\Psi_{\infty-1} = \bigcup_{i=1}^{\infty} \Psi^i, \quad \Psi^i = \prod_{k \neq i} \bar{I}_k. \quad (21)$$

Recall proposition (2.2) where we associated a unique function  $h_{\xi}$  with  $\Sigma$  and  $\xi \in \Psi$ . This collection of functions will be denoted as,

$$\mathcal{H} = \{h_{\xi} : \xi \in \Psi\}. \quad (22)$$

Given a point  $\xi \in \Psi$ , we can write it as  $(\xi_i, \xi^i)$  with  $\xi_i \in \bar{I}_i$  and  $\xi^i \in \Psi^i$ . We then consider the analytic function in  $\Pi$  given by,

$$\omega_{\xi^i} = \frac{1}{\lambda - \xi_i} h_{\xi}. \quad (23)$$

It is not hard to see from the product representation for  $h_{\xi}$  in  $\Pi$  that  $\omega_{\xi^i}$  is independent of  $\xi_i \in \bar{I}_i$ . This class of functions has the following properties in the gaps. For the following proposition, we set the distances of  $\tau_0$  and  $\tau_{\infty}$  from  $I_k$  to be  $d_{0k}$  and  $d_{\infty k}$ .

**Lemma 3.2.** Consider the  $\omega_{\eta}$  for  $\eta \in \Psi_{\infty-1}$ . Then it has the following properties. There is an  $i \in \mathbb{Z}^+$  such that:

1.  $\omega_{\eta}(x) > 0 \quad x \in I_i$ .
2.  $\omega_{\eta}(x) > 0 \quad x \in [\cup_{I_k < I_i} (\tau_{2k-1}, \eta_k)] \cup [\cup_{I_k > I_i} (\eta_k, \tau_{2k})]$ .
3.  $\omega_{\eta}(x) < 0 \quad x \in I(\infty) \cup [\cup_{I_k > I_i} (\tau_{2k-1}, \eta_k)] \cup [\cup_{I_k < I_i} (\eta_k, \tau_{2k})] \cup I(-\infty)$ .
4.  $\omega_{\eta}(\lambda) = O\left(\frac{1}{\lambda^2}\right), \quad \lambda \rightarrow \infty$
5. The following bounds are valid in  $\bar{I}_k$  for each  $k \neq i$ ,

$$\omega_{\eta}(\lambda) \leq \left| \frac{q_k}{\sqrt{d_{0k}d_{\infty k}}} \frac{\prod_{j \neq k} (1 + (l_j/s_j))}{\sqrt{(\lambda - \tau_{2k-1})(\lambda - \tau_{2k})}} \right|$$

and for  $k = i$ ,

$$\left| \frac{\prod_{j \neq i} (1 - (l_j/s_j))}{\sqrt{(\lambda - \tau_{2i-1})(\lambda - \tau_{2i})}} \omega_{\eta}(\lambda) \right| \leq \left| \frac{\prod_{j \neq i} (1 + (l_j/s_j))}{\sqrt{(\lambda - \tau_{2i-1})(\lambda - \tau_{2i})}} \right|.$$

*Proof.* Since  $\eta$  belongs to  $\Psi_{\infty-1}$ , which is a disjoint union of  $\Psi^k$ 's, there is an  $i \in \mathbb{Z}^+$  with  $\eta \in \Psi^i$ . We consider any point  $\xi_i \in I_i$  and let  $\xi = (\xi_i, \eta)$ . Then  $\xi \in \Psi$  and we have a  $h_\xi$  associated with this  $\xi$ . Using this we define the  $\omega_\eta$  as in equation (23). Then the properties (1)–(3) stated for  $\omega_\eta$  are clear from those of  $h_\xi$ , from proposition (2.2) and the fact that  $1/(\lambda - \xi_i)$  is negative in  $(-\infty, \xi_i)$  and positive in  $(\xi_i, \infty)$ . The asymptotic behaviour of  $\omega_\eta$  is clear from that of  $h_\xi$ . Given the asymptotic behaviour and the analyticity of  $\omega_\eta$  in  $I(-\infty)$  and  $I(\infty)$ , it is integrable outside the compact set  $[a, b]$  with  $a < \tau_0$  and  $b > \tau_\infty$ . Therefore we consider a compact set  $[a, b]$  and show the integrability there. As before we associate a  $\xi \in \Psi$  with  $\eta$  with  $\xi_i$  chosen in (for example the mid point of the gap)  $I_i$ . We write the product representation of (12) for  $h_\xi$  and write it as

$$h_\xi = \frac{(\lambda - \xi_k)}{\sqrt{(\lambda - \tau_{2k-1})(\lambda - \tau_{2k})}} g_k(\lambda) \quad (25)$$

with

$$g_k(\lambda) = \frac{\sqrt{(\lambda - \tau_{2k-1})(\lambda - \tau_{2k})}}{(\lambda - \xi_k)} h_\xi(\lambda).$$

Then in the expression for  $g_k(\lambda)$ , the numbers  $\tau_{2k-1}$ ,  $\tau_{2k}$ ,  $\xi_k$  do not occur in the numerator or denominator. So for each  $\lambda \in I_k$ ,  $k \neq i$ , the  $\text{dist}(\lambda, I_j) > s_j$ , so that we use the bounds of lemma (3.1) with  $d_j = s_j$ , and  $d_0 = d_{0k}$  and  $d_\infty = d_{\infty k}$  both of which are bigger than  $s_k$  by assumption (2), to conclude that the bounds,

$$\omega_\eta(\lambda) \leq \left| \frac{1}{\sqrt{d_{0k} d_{\infty k}}} \frac{l_k}{s_k} \frac{\prod_{j \neq k} (1 + (l_j/s_j))}{\sqrt{(\lambda - \tau_{2k-1})(\lambda - \tau_{2k})}} \right|$$

are valid for  $\lambda \in I_k$ ,  $k \neq i$ . As for the case  $k = i$ , the bound

$$\omega_\eta(\lambda) \leq \left| \frac{1}{d_{0i} d_{\infty i}} \frac{\prod_{j \neq i} (1 + (l_j/s_j))}{\sqrt{(\lambda - \tau_{2i-1})(\lambda - \tau_{2i})}} \right|$$

is clear, since the factors  $(\lambda - \xi_i)$  occurring in the numerator of the product representation of  $h_\xi$  and that occurring in the denominator, in the definition of  $\omega_\eta$ , cancel. The lower bounds of (5) are deduced similarly, by noting that the distance of a point in  $I_i$  to either of  $\tau_0$ ,  $\tau_\infty$  is bounded above by  $\tau_\infty - \tau_0$ . The infinite products occurring in the numerators of the inequalities in (5) converge by the summability of  $q_i$  of assumptions (2). The integrability in the closure of the gaps follows from (5). In the case of  $\lambda$  in  $\Sigma$ , under the assumptions (2) and theorem (A.2.3) the limits  $h_\xi(\lambda + i0)$  exist everywhere in the interior of  $\Sigma$  and are purely imaginary there by construction of  $h$ . Therefore the absolute value of  $h_\xi(\lambda + i0)$  is just the density of the absolutely continuous measure (or a constant multiple of it) representing  $h_\xi$  hence it is integrable once we note that the factor  $1/(\lambda - \xi_i)$  is uniformly bounded in  $\lambda \in \Sigma$  for each fixed  $i$ , by the choice of  $\xi_i$ .

We consider the real Banach space, with  $q_i$  as in assumption (2) and  $\delta_i$  denoting the point measure at  $i$  with unit mass,

$$\mathcal{X} = l_{\mathbb{R}}^\infty(\mathbb{Z}^+, \sigma), \quad \sigma = \sum_{i=1}^{\infty} q_i^{-1} \delta_i$$

Using this Banach space and the set  $\Psi_{\infty-1}$  we can define a class of functions  $\mathcal{H}_{\infty-1}$  analytic in  $\Pi$  as,

$$\mathcal{H}_{\infty-1} = \left\{ \sum_{i=1}^{\infty} \kappa_i \omega_{\xi^i} : \xi^i \in \Psi^i \text{ and } \kappa \in \mathcal{K} \right\}. \quad (26)$$

We will use the functions in  $\mathcal{H}_{\infty-1}$  to construct a class of differentials of the first kind on a Riemann surface to be considered in the next section. The properties of functions in  $\mathcal{H}_{\infty-1}$  will also be important for the properties of the Abel–Jacobi map that will be constructed later. Therefore we start with some of the simple properties. To do this we need a technical lemma.

**Lemma 3.3** *Given any  $N$ , we have finitely many, say  $M(N)$ , non intersecting rectangular curves  $R_N^i$ ,  $i = 1, \dots, M(N)$  with the following properties.*

1. Each  $R_N^i$  has its sides parallel and perpendicular to the axes and each side parallel to the  $y$ -axis goes through the mid point of some band in  $\Sigma$ .
2. Every point of  $\mathcal{E}$  is enclosed by some  $R_N^i$ .
3. There is a sequence of positive numbers  $m_N$  increasing to infinity such that the sum of the perimeters,  $\text{Per}(R_N^i)$ , of the curves  $R_N^i$  satisfies,

$$\sum_{i=1}^{M(N)} \text{Per}(R_N^i) < \frac{1}{m_N}.$$

*Proof.* Suppose the cardinality of the set  $\mathcal{E}_1$  is  $l$  and let the minimum of the distance between the points of  $\mathcal{E}_1$  be  $d$ , which is positive since the points are distinct. We fix an  $N$  and consider the closed intervals  $J_N^i$ ,  $i = 1, \dots, l$  of positive length smaller than  $d/12^{N+1}$  with each point  $p_i \in \mathcal{E}_1$  as its mid point and such that the end points of the intervals do not coincide with any point of  $\mathcal{E}$ . This choice is possible since the points in  $\mathcal{E}_1$  are finite. Then it is clear that the number of points of  $\mathcal{E}$  in  $[\tau_0, \tau_\infty] \setminus \cup_i J_N^i$  is finite. Let this number be  $M(N)$ . Let the minimum of their distance be  $d(M(N))$ . We now pick intervals  $J_N^j$ ,  $j = l+1, \dots, M(N)+l$ , of length equal to  $d(M(N))/(M(N)2^{N+1})$  with these points as their mid points. By our choice, the end points of these intervals do not coincide with any points of  $\mathcal{E}$ . The choice of the rectangular curves is made as follows. Consider a point  $p_j \in \mathcal{E}_1 \cup [\mathcal{E} \setminus \cup_i J_N^i]$ ,  $j = 1, \dots, M(N)+l$ . Then by assumptions (2) it follows that there are bands  $b_j^1$  and  $b_j^2$ , to the left and right of  $p_j$  respectively contained in  $J_N^j$ . It is also clear, since the endpoints of  $J_N^i$ ,  $i = 1, \dots, l$  are not points of  $\mathcal{E}$ , we can choose  $b_j^1$ ,  $b_j^2$  such that the points of  $\mathcal{E} \cap J_N^i$  are to the left of  $b_j^2$  and to the right of  $b_j^1$ . For each  $j$ , now, we choose the rectangular curve  $R_N^j$  with sides parallel and perpendicular to the axes such that, the sides parallel to the  $y$ -axis pass through the mid points of  $b_j^1$  and  $b_j^2$ . The sides parallel to the  $x$ -axis are at a distance equal to maximum of  $\text{dist}(p_j, b_j^k)$ ,  $k = 1, 2$ . Clearly the perimeters of  $R_N^j$  satisfy the bounds  $\text{Per}(R_N^j) \leq \text{constant } 1/(12^{N+1})$  for  $j = 1, \dots, l$  and  $\text{Per}(R_N^j) \leq \text{constant } 1/(M(N)2^{N+1})$  for  $j = l+1, \dots, M(N)+l$ . The properties listed in the lemma are clear with the redefinition of  $M(N)+l$  as  $M(N)$  and  $m_N = \text{constant } 2^{N+1}$ .  $\square$

For later use we shall consider, a different set  $r_N$  of rectangles for each  $N$  given by  $\cup_{i=1}^{M(N)} r_N^i$ . Where  $r_N^i$  is a rectangle with its sides parallel to the  $x$ - and  $y$ -axes, the distance

of the sides parallel to the  $x$ -axis being  $s_0$ , one of the sides parallel to the  $y$ -axis passes through  $\tau_0$  and the other side contains a side of  $R_N^i$  given in the above lemma.

We first show an interpolation theorem for  $\omega_\eta$ 's, for which recall the definition of  $\xi \in \Psi_{\infty-1}$  associated with  $\xi \in \Psi$  and the definition of the set  $S_\xi$  given in the last section.

### PROPOSITION 3.4

Consider  $\eta \in \Psi_{\infty-1}$ , and consider  $\omega_\eta$ . Then there is an  $i \in \mathbb{Z}^+$  such that for any  $\xi \in \Psi$ , we have the following relation as analytic functions in  $\Pi$ .

$$\omega_\eta(\lambda) = h_\xi(\lambda) \sum_{k=1}^{\infty} \frac{D_k^i}{(\lambda - \xi_k)} \quad (27)$$

where

$$D_k^i = \left( \frac{\xi_k - \eta_k}{\xi_k - \xi_i} \right) \prod_{j \neq k, i} \frac{(\xi_k - \eta_j)}{(\xi_k - \xi_j)} \quad D_i^i = \prod_{j \neq i} \frac{(\xi_i - \eta_j)}{(\xi_i - \xi_j)}. \quad (28)$$

with the sum converging compact uniformly in  $\mathbb{C} \setminus \mathcal{E} \cup S_\xi$ .

*Proof.* Since  $\eta$  is in  $\Psi_{\infty-1}$ , there is an  $i$  such that  $\eta \in \Psi^i$ . This is the  $i$  stated in the proposition. So we fix the  $i$  and consider the rational function  $f(\lambda) = \omega_\eta(\lambda)/h_\xi(\lambda)$  in  $\Pi$ . By using the product representation for  $h_\xi$  and the definition of  $\omega_\eta$  we see that formally,

$$f(\lambda) = \frac{1}{(\lambda - \xi_i)} \prod_{j \neq i} \frac{(\lambda - \eta_j)}{(\lambda - \xi_j)}.$$

By using the estimates

$$\left| \frac{(\lambda - \eta_j)}{(\lambda - \xi_j)} \right| \leq 1 + \left| \frac{(\xi_j - \eta_j)}{(\lambda - \xi_j)} \right| \leq 1 + \left| \frac{l_j}{(\lambda - \xi_j)} \right|. \quad (29)$$

and the convergence of the products  $\prod (1 + q_i)$  a computation shows that the product defining  $f(\lambda)$  converges uniformly in compacts of  $\mathbb{C} \setminus (\mathcal{E} \cup S_\xi)$  and it is also clear that the points  $\xi_k$  are simple poles of  $f(\lambda)$  in  $\mathbb{C} \setminus \mathcal{E}$ . Thus showing that  $f(\lambda)$  extends to a meromorphic function, with simple poles at  $\xi_k$ 's, in  $\mathbb{C} \setminus \mathcal{E}$ . The residues of the function  $f(\lambda)$  at the poles  $\xi_k$  when computed are precisely the quantities  $D_k^i$  stated in the proposition.

On the other hand the function

$$g(\lambda) = f(\lambda) - \sum_{j=1}^{\infty} \frac{D_j^i}{(\lambda - \xi_j)}$$

is analytic in  $\mathbb{C} \setminus \mathcal{E}$  and the bounds of (29) together with the bounds  $(\lambda - \xi_i) > s_j$ , which is valid in view of assumption (2), for  $\lambda$  on the rectangles, shows that we have

$$\sup_{\lambda \in R_N^i} g(\lambda) \leq \prod (1 + q_j) + \sum_{k=1}^{\infty} \prod_{j \neq k, i} (1 + q_j) \frac{l_k}{s_k}$$

where we have used the estimates for  $D_k^i$  from lemma (3.5) which will be shown below.

The right hand side in the above inequality is independent of  $N, l$  etc., so that from the definition of the rectangles and the proposition (3.3), the above uniform bound on them shows that the points in  $\mathcal{E} \setminus \mathcal{E}_1$  are not essential singularities of  $g(\lambda)$ . The same bound also shows that

$$\lim_{N \rightarrow \infty} \int_{\lambda \in \bigcup_{i=1}^{M(N)} R_N^i} (\lambda - x)^m g(\lambda) = 0$$

for each positive integer  $m$ , and hence in the Laurent series expansion of  $g$  at the points  $x \in (\mathcal{E} \setminus \mathcal{E}_1)$  all the negative coefficients vanish, showing that  $g$  can be extended as an analytic function to  $\mathbb{C} \setminus \mathcal{E}_1$ . A similar argument shows that  $g$  can be extended to an analytic function in  $\mathbb{C}$ . But  $g$  has zeros at  $\xi_i$  and these accumulate to some point in  $\mathcal{E}$ , its region of analyticity, showing that  $g$  has to be identically zero.  $\square$

**Lemma 3.5.** Consider  $\eta \in \Psi^i$  and  $\xi \in \Psi$ . Let  $D_k^i$  be as in proposition (3.4). Then the following bounds are valid for  $k \neq i$ ,

$$|D_k^i| \leq \left| \frac{\xi_k - \eta_k}{\xi_k - \xi_i} \prod_{j \neq k, i} (1 + q_j) \right| \leq C \min \left( q_k, \frac{l_k}{s_i} \right), \quad C_1 \leq |D_i^i| \leq C_2$$

with  $C_1, C_2$  independent of  $i$ .

*Proof.* From the definition of  $D_k^i$  it is clear that we can write

$$D_k^i = \frac{\lambda - \eta_k}{\lambda - \xi_i} \prod_{j \neq k, i} \frac{\lambda - \eta_j}{\lambda - \xi_j} \Big|_{\lambda = \xi_k}$$

which shows using the bounds of (29) with  $\lambda = \xi_k$ , and the lower bounds  $\xi_k - \xi_i \geq \min(s_i, s_k)$  implied by assumption (2). The uniform upper and lower bounds, in the case of  $k = i$  are obvious from the expressions.  $\square$

**Theorem 3.6.** (Interpolation theorem). Consider  $f$  in  $\mathcal{H}_{\infty-1}$  and let  $\xi \in \Psi$ . Then we have the following relation as analytic functions in  $\Pi$ .

$$f(\lambda) = h_\xi(\lambda) \sum_{i=1}^{\infty} \frac{C_i f(\xi_i)}{(\lambda - \xi_i)}, \quad C_i = \frac{\sqrt{R(\xi_i)}}{\prod_{j \neq i} (\xi_i - \xi_j)}. \quad (30)$$

In particular  $f$  can be written as

$$f(\lambda) = \sum_{i=1}^{\infty} \kappa_i(f) \omega_{\xi_i}, \quad \kappa(f) \in \mathcal{K} \quad (31)$$

*Proof.* We consider a function  $f \in \mathcal{H}_{\infty-1}$  given by

$$f(\lambda) = \sum_{i=1}^{\infty} \kappa_i \omega_{\xi_i}, \quad \kappa \in \mathcal{K}$$

with  $\xi^i \in \Psi^i$ . Then using the proposition (3.4) we see that each of the  $\omega_{\xi_i}$ 's can be

written as

$$\omega_{\zeta_i}(\lambda) = h_{\zeta}(\lambda) \sum_{k=1}^{\infty} \frac{D_k^i}{(\lambda - \xi_k)}.$$

From this the theorem follows after an interchange of sum and the relation  $C_k f(\xi_k) = \sum_{i=1}^{\infty} D_k^i \kappa_i$ . The necessary convergences can be checked using the bounds

$$|\kappa_k(f)| = |C_k f(\xi_k)| = \sum_{i=1}^{\infty} |D_k^i \kappa_i| \leq |D_k^k \kappa_k| + \left| \sum_{i \neq k}^{\infty} \kappa_i D_k^i \right| \leq C q_k$$

which follow from the bounds in lemma (3.5) and from the definition of  $\mathcal{H}$ .  $\square$

Since any two elements of  $\mathcal{H}_{\infty-1}$  can be written in terms of a single  $h_{\zeta}$ ,  $\zeta \in \Psi$ , addition of two elements, via the above theorem gives again an element of  $\mathcal{H}_{\infty-1}$ . Therefore  $\mathcal{H}_{\infty-1}$  is a linear space, in fact a vector space over the reals. This fact will be useful for the following corollary.

### COROLLARY 3.7

Suppose  $f$  is in  $\mathcal{H}_{\infty-1}$  such that it vanishes at a point each in each of the gaps. Then  $f \equiv 0$ .

*Proof.* Consider the point of  $\zeta \in \Psi$  such that  $S_{\zeta}$  is the set of zeros of  $f$ , in the gaps. Then, by the interpolation theorem, we can write  $f$  in terms of  $h_{\zeta}$ , and the numbers  $f(\xi_i)$ , as in (30), so that  $f$  is identically zero.  $\square$

In the following we choose a basis for  $\mathcal{H}_{\infty-1}$  which will come useful for the analysis on the Riemann surface and in terms of which the Abel–Jacobi map will be defined later.

### PROPOSITION 3.8

There exists a  $\zeta \in \Psi$  and a collection  $\{\sigma_i\}$  of positive numbers  $C \leq \sigma_i \leq D$  such that  $f_i(\lambda) = \sigma_i \omega_{\zeta_i}$  have the following properties

1.  $\int_{I_j} |f_i(x)| dx < \infty \quad \forall j$
2.  $\int_{I_j} f_i(x) dx = 0 \quad \forall i \neq j$
3.  $\int_{I_i} f_i(x) dx = \frac{\pi}{2}.$

*Proof.* The first property mentioned in the definition is automatic for  $f_i$  defined with any  $\zeta \in \Psi$ , from the estimates (5) of lemma (3.2). For the second we consider an  $i$  and fix it. Then we consider a positive integer  $n$  and consider the following subset of  $\Psi^i$

$$\Lambda(n) = \prod_{i=1}^n \bar{I}_i \prod_{j=1}^n \{\tau_{i,j}\}$$

are strictly positive in the gaps  $I_j$ ,  $j = i$  and  $j > n+1$  and they are real in  $I_j$  for the remaining values of  $j$ . Further it is also clear, from lemma (3.2) that whenever  $1 \leq j \leq n+1$ ,  $j \neq i$ , we have  $\omega_{\xi^1}(\lambda) > 0$  or  $< 0$  according as  $\lambda > \xi_j^i$  or  $< \xi_j^i$  in  $I_j$ . In the proof below we take  $i = 1$ , however the proof works for any  $i$ . (For  $i$  not equal to 1 it may be necessary to multiply  $F^n$ s by  $\pm 1$ ). We consider the following function  $F^n$  of  $n$  real variables.

$$F_j^n(x_2, \dots, x_{n+1}) = \int_{I_j} \omega_{\xi^1}(\lambda) d\lambda \quad j = 2, \dots, n+1$$

with  $x_j = \xi_j^1$   $j = 2, \dots, n+1$ . Then from the properties of  $\omega_{\xi^1}$ , when the first  $n$  coordinates of  $\xi^1$  vary, we find that

$$F_j^n(x_2, \dots, \tau_{2k-1}, \dots, x_{n+1}) < 0 \quad k = 2, \dots, n+1$$

and

$$F_j^n(x_2, \dots, \tau_{2k}, \dots, x_{n+1}) > 0, \quad k = 2, \dots, n+1.$$

It is also clear that  $F^n$  is a continuous function from  $X_{j=2}^{n+1} \bar{I}_j$  to  $\mathbb{R}^n$  satisfying the assumptions of theorem (A.3), so that the existence of a point  $\xi^1(n)$  in  $\Psi^i$  at which  $F^n$  vanishes is guaranteed. When  $\Psi^1$  is equipped with the topology of pointwise convergence, the subsequence of the points  $\xi^1(n)$  converges, to say  $\xi^1$ , in  $\Psi^1$ . Under this convergence, we have the pointwise convergence of  $\omega_{\xi^1(n)}$  as an integrable function on each  $I_j$ ,  $j = 1, 2, \dots$ . It is also clear from the estimates (5) of lemma (3.2), valid for any  $\eta \in \Psi_{\infty-1}$ , that  $\omega_{\xi^1(n)}$  are integrable uniformly in each  $I_k$ ,  $k = 1, 2, \dots$ . Thus by Lebesgue dominated convergence theorem we have

$$\int_{I_k} \omega_{\xi^1}(\lambda) d\lambda = 0 \quad \forall k \neq i.$$

It is also, clear that for  $k = 1$ , the above integral is nonzero and in fact positive hence the normalization is a matter of choosing an overall real multiplicative constant  $\sigma_1$ . For later purposes, we shall write the functions  $f_i$  as

$$f_i(\lambda) = \sigma_i \omega_{\xi^i}(\lambda), \quad \text{with} \quad C \leq \sigma_i \leq D. \quad (32)$$

The upper and lower bounds for  $\sigma_i$  follow from the estimates (5) of lemma (3.2) and assumption (2.1), which also show that the constants  $C$  and  $D$  are independent of  $i$ .  $\square$

### PROPOSITION 3.9

*Assume that  $\Sigma$  satisfies assumptions (2). Assume further that the measure  $\mu$  in the Herglotz representation of proposition (2.2) is absolutely continuous for each  $\xi \in \Psi$ . Then there exists a  $\zeta$  in  $\Psi$  such that  $h_\zeta$  has the following properties*

1.  $\int_{I_j} h_\zeta(x) dx = 0 \quad \forall j$
2.  $\text{Re} \int_{\Sigma \cap (-\infty, \tau_{2i})} h_\zeta(x) dx = 0 \quad \forall i$ .

*Proof.* Given any  $\xi \in \Psi$  we construct  $h_\xi$  as in proposition (2.2). By the assumption on  $\Sigma$ , the limits  $h_\xi(\lambda + i0)$  of the functions  $h_\xi$  of proposition (2.2) exist everywhere in the



interior of  $\Sigma$  and by assumption there is no singular part for  $\mu$  so that  $d\mu(x) = \text{Im } h_\xi(x + i0)dx$ . The product representation for  $h_\xi$  also shows, using estimates of lemma (3.1), that  $h_\xi(\lambda)$  is integrable in each  $I_i$ . The proof of the first item follows the proof of the previous proposition, by which we make a choice of a fixed point  $\zeta \in \Psi$ , so we omit it. The second item follows since  $\text{Re } h_\xi$  is zero almost everywhere in the spectrum.  $\square$

There is an interesting corollary of proposition (3.8) which will be useful in showing that the Abel–Jacobi map to be introduced later is onto. In the following we use the usual convention that if  $x < y$ , then  $\int_y^x f(\lambda) = - \int_x^y f(\lambda)$ .

#### COROLLARY 3.10

Consider two points  $\zeta$  and  $\eta$  in  $\Psi$ . If for all  $f \in \mathcal{H}_{\infty-1}$ ,

$$\sum_{j=1}^{\infty} \int_{\eta_j}^{\zeta_j} f(\lambda) d\lambda = 0$$

then  $\zeta \equiv \eta$ .

*Proof.* Consider the  $\xi^1$  in  $\Psi^1$  given by  $\xi_j^1 = \eta_j$ ,  $j \neq 1$ . With this choice we see, from properties (1, 2, 3) of lemma (3.2), that  $\omega_{\xi^1} > 0$  in  $(\eta_j, \zeta_j)$  whenever  $\eta_j < \zeta_j$ , and  $\omega_{\xi^1} < 0$  in  $(\zeta_j, \eta_j)$  whenever  $\zeta_j < \eta_j$  for any  $j \neq 1$  and  $\omega^1 > 0$  in  $I_1$ . Therefore the convention mentioned before the corollary implies that for each  $j = 1, 2, \dots$  we have

$$\int_{\eta_j}^{\zeta_j} \omega_{\xi^1} d\lambda \geq 0.$$

Since  $\omega_{\xi^1} \in \mathcal{H}_{\infty-1}$  the corollary is immediate.  $\square$

### 4. Analysis on a Riemann surface

We consider a Riemann surface associated with the set  $\Sigma$  of the last section and consider the Abel–Jacobi map associated with the surface. We compute the image of a class of divisors on the surface under the Abel–Jacobi map. We make some assumptions on the divisors, which will be verified in § 5 for the application to almost periodicity of a sequence of points on the Riemann surface. In a subsection we reformulate the Abel–Jacobi map on a real Banach space and show that the map and its inverse are Lipschitz continuous bijections of the Banach space to itself with some regularity properties. These results will also be needed in the proof of almost periodicity of the Jacobi matrix.

#### 4.1 The Riemann surface

In this section we construct a Riemann surface  $\mathcal{R}$  whose branch points are the  $\tau_i$ 's and the  $\tau_{\infty}, \tau_0$ 's. We take two copies of the Riemann sphere and delete the points of  $\mathcal{E}$  from each. The rest of the procedure to construct  $\mathcal{R}$  is similar to the construction of a hyperelliptic Riemann surface with finite number of branch points. On each

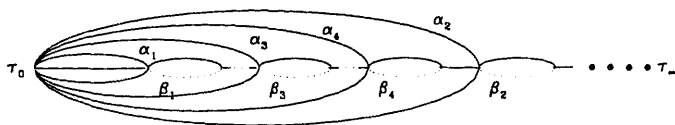


Figure 1.  $\alpha$  and  $\beta$  cycles. Dashed lines indicate change of sheet.

sphere we cut the real line along the spectral bands and glue the resulting spheres together by joining the upper lips of the cuts of the first one to the corresponding lower lips of the cuts of the second one and vice versa. Thus we get a non-compact hyperelliptic surface of infinite genus. This surface will be a two sheeted branched cover of the Riemann sphere with the points of  $\mathcal{E}$  removed from it.

We denote by  $\lambda$  the projection map from  $\mathcal{R}$  into the Riemann sphere. On  $\mathcal{R}$  we give a coordinate system as follows. We denote the points of  $\lambda^{-1}(\infty)$  in the upper and lower sheets as  $\infty_1$  and  $\infty_2$ . At  $\infty_1$  we give local coordinates by mapping a neighbourhood of it to a neighbourhood of the origin in  $\mathbb{C}$  by the map  $z(\infty_1) = 0$  and  $z(p) = 1/\lambda(p)$  for other points  $p$  in the neighbourhood. Similarly we give coordinates around  $\infty_2$ . At point  $p_n$  corresponding to a branch point  $\tau_n$  we give the coordinate system given by  $z^2 = \lambda(p) - \tau_n$ . For the rest of the points we give coordinate systems via the projection map  $\lambda$ . Thus the Riemann surface is a two sheeted branched cover of the Riemann sphere with  $\{\tau_0, \tau_\infty, \tau_{2i}$  and  $\tau_{2i+1}\}$   $i = 1, 2, \dots$  as the branch points.

We now give some definitions, see figure 1.

#### DEFINITION 4.1

The closed loop which starts from  $\tau_{2i-1}$ , goes to  $\tau_{2i}$  through the embedded spectral gap on the lower sheet and comes back to  $\tau_{2i-1}$  through the similar spectral gap on the upper sheet is called the  $i$ -th  $\beta$ -cycle and is denoted by  $\beta_i$ .

Let  $P_0$  be any point on the embedded real line on the lower sheet to the left of  $\tau_0$ . The closed loop, lying on the lower sheet, that starts from some point on  $\beta_i$  going to  $P_0$  and coming back to the same point through the other side of the lower sheet is called the  $i$ -th alpha cycle and is denoted by  $\alpha_i$ .

We recall that a differential of the first kind on  $\mathcal{R}$  is one which can be written in local coordinates  $z$  as  $g(z)dz$  with  $g$  holomorphic.

We shall consider a class of differentials on  $\mathcal{R}$  which will be normalized on using the  $\alpha$  and  $\beta$  cycles considered above.

**Lemma 4.2.** *Each of the functions of class  $\mathcal{H}_{\infty-1}$  considered in the last section gives rise to a differential of the first kind on  $\mathcal{R}$ .*

*Proof.* Consider the differentials  $f(\lambda)d\lambda$  on  $\mathcal{R}$  with  $f \in \mathcal{H}_{\infty-1}$ . A priori this is a differential defined on  $\lambda^{-1}(\Pi)$  and a computation in each of the local coordinate charts on  $\mathcal{R}$  shows that it extends to a differential of the first kind on the whole of  $\mathcal{R}$ .  $\square$

We fix a basis of the differentials of the first kind on  $\mathcal{R}$ . Recall the functions  $f_i$  of proposition (3.8). Using these we form the following differentials of the first kind on

given by

$$d\omega_i(p) = f_i(\lambda(p))d\lambda(p)$$

where the right hand side is written in local coordinates at  $p$  in  $\mathcal{R}$ .

Then the normalizations of proposition (3.8) immediately imply on computation of the product representation for  $f_i$ 's, the choice of the  $\alpha$  and  $\beta$  cycles, that

$$\int_{\beta_j} d\omega_i = 2 \int_{I_j} f_i(\lambda) d\lambda = \delta_{ij} \pi \quad (33)$$

$$\int_{\alpha_j} d\omega_i = 2 \int_{(-\infty, \tau_{2i-1}) \cap \Sigma} f_i(\lambda) d\lambda = \pi_{ij}$$

where  $\pi_{ij}$  is purely imaginary.

Connected with the basis given above and a fixed point  $p_0$  in  $\mathcal{R}$  we define the Abel map

$$\omega_i(p) = \int_{p_0}^p d\omega_i \quad (34)$$

which is a function from  $\mathcal{R}$  to  $\mathbb{C}$  for each  $i$ . It is not well defined since the right hand side depends on the path of integration. It is however well defined as a map into the Jacobian variety of  $\mathcal{R}$ . We shall not go into this further. We shall use the Abel map for defining the Abel–Jacobi map later.

The above collection of differentials of the first kind form indeed a basis for a real Hilbert space of holomorphic differentials of the first kind. It is possible to define and analyze the period matrix, prove the bilinear relations of Riemann and define the theta functions on it. However, we shall not discuss these points in this work. For the present purposes it is enough to quickly get a consequence of an analogue of the bilinear relations of Riemann. More precisely we are interested in the image of the Abel–Jacobi map on a class of divisors. To present this we consider the subset  $\mathcal{R}$  coming from

$$\Phi = \lambda^{-1}(\Psi) \cup \lambda^{-1}(\infty)$$

where  $\Psi$  is as in § 2.

We recall that a divisor is a formal sum

$$D = \sum_{P \in \mathcal{R}} D(P)P$$

with  $D(P)$  an integer valued function on  $\mathcal{R}$ . We shall denote by the support of a divisor as the set of points  $P$  with  $D(P)$  non-zero. We recall that the divisor  $D_f$  of a meromorphic function  $f$  is  $D_f(P) = \text{ord}_P(f)$ . A divisor  $D$  is said to be principal if it is the divisor of a meromorphic function. We consider below a special subclass of principal divisors  $D$  given by

$$D = K[P_\infty - Q_\infty] + \sum_{i=1}^{\infty} P_i - Q_i \quad (35)$$

for  $P_i, Q_i \in \Phi$  for an integer  $K$ .

The divisors of the above type will occur as the divisors of some  $m$ -functions of random Jacobi matrices which will be considered in the last section.

We shall denote the two sheets of  $\mathcal{R}$  by  $\mathcal{R}^\pm$  in the analysis below. We consider the rectangular curves introduced in the last section in lemma (3.3), and let

$$S_N^j = \lambda^{-1}(R_N^j) = \cup_{k=1}^2 S_N^j(k), \quad 1 \leq j \leq M(N) \quad \forall N.$$

Since each point of  $R_N^j$  is away from the branch points,  $R_N^j$  has two preimages in  $\mathcal{R}$  and these will be disjoint, (see figure 2 to get an idea of the preimages, where dotted lines indicate change of sheets) so that the above union is a disjoint union. We denote the closed region in  $\mathbb{C}$  bounded and enclosed by the rectangles  $R_N^j$  of lemma (3.3) by  $\bar{R}_N^j$  and let

$$\bar{S}_N^j = \lambda^{-1}(\bar{R}_N^j), \quad 1 \leq j \leq M(N) \quad \forall N.$$

We first note that since the Riemann surface we have is of infinite genus, we need to do an approximation to obtain the relation (36). To this end we consider the open region  $\mathcal{R}_N$  of  $\mathcal{R}$  obtained by taking the sets  $\bar{S}_N^k$ , defined above (if necessary with smoothed out to make the boundary smooth)

$$\mathcal{R}_N = \mathcal{R} \setminus \bigcup_{i=1}^{M(N)} \bar{S}_N^i.$$

Then  $\mathcal{R}_N$  is increasing with  $N$  in the set theoretic sense. Further,  $\mathcal{R}_N$  is a Riemann surface by its own right. It can be compactified by adding  $2M(N)$  disks. Therefore if we fix a collection of  $\alpha$  and  $\beta$  cycles in  $\mathcal{R}_N$ , there is a normal polygon  $D_N$ , (we refer to [10] for a construction of this domain) with the number of sides equal to 4 times the number of beta cycles in  $\mathcal{R}_N$  and having  $2M(N)$  holes corresponding to the removal of  $M(N)$  disjoint closed sets  $\bar{S}_N^i$ . (We should remember here that each of  $\bar{S}_N^i$  is a closed subset of  $\mathcal{R}$  having two preimages for each point of  $\bar{R}_N^i$  except the branch points.) Suppose that the number of beta cycles in  $\mathcal{R}_N$  is  $k(N)$ . Then the  $4k(N)$  sides of the polygon are indexed as  $a_i$  or  $b_i$  according to whether they correspond to traversing an alpha or a beta cycle in the anti-clockwise or as  $a_i^{-1}$  or  $b_i^{-1}$  corresponding to traversing them in the clockwise direction. Further the sides are such that  $a_i$ ,  $b_i$ ,  $a_i^{-1}$  and  $b_i^{-1}$  occur as consecutive sides of the polygon for each  $i$ . There are  $2M(N)$  holes  $H_j$  in  $D_N$  we shall index them so that  $\partial H_{2j-1}$  and  $\partial H_{2j}$ ,  $j = 1, 2, \dots, M(N)$ , are  $S_N^j(1)$  and  $S_N^j(2)$  respectively.

We find that the Abel map  $\omega_i$  in  $D_N$  is not single valued, though it has analytic continuation along any path in  $D_N$ , the values along two different paths differing by an integer multiple of its integrals on the boundaries of the holes. Therefore to obtain analytic functions corresponding to the Abel map, which essentially means that we ensure that the paths of integration do not wind around the holes, we make cuts in  $D_N$  along paths  $\gamma_{2j-1}$ ,  $\gamma_{2j}$  connecting a fixed vertex to the boundary of the holes  $\partial H_{2j-1}$ ,  $\partial H_{2j}$ , so that we obtain a simply connected domain which we call  $E_N$ . Recall

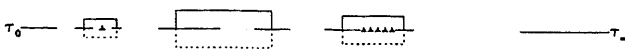


Figure 2. The rectangles  $R_N^j$ .

the set  $r_N$  of rectangles defined after lemma (3.3). For example the  $\gamma_{2j-1}$ ,  $\gamma_{2j}$  can be chosen to be the arcs  $r_N^j(1)$  and  $r_N^j(2)$  respectively given by the portions of  $\lambda^{-1}(r_N^i)$  in  $\lambda^{-1}(\mathbb{C}^\pm) \cap \mathcal{R}^+$  for the purpose of estimates needed later.

### PROPOSITION 4.3

Consider two points  $P, Q \in \Phi$  and suppose  $\phi$  is a meromorphic function on  $\mathcal{R}$  such that it has simple poles at  $Q_i$ , simple zeros at  $P_i$   $i = 1, 2, \dots$ , and has a zero and a pole of order  $K$  at  $P_\infty$  and  $Q_\infty$  respectively. Suppose further  $\phi$  satisfies the conditions that,

$$\lim_{N \rightarrow \infty} \sum_{j=1}^{M(N)} \left| \int_{\partial H_j} \frac{d\phi}{\phi} \right| + \left| \int_{\partial H_j} \omega_i \right| = 0 \quad \sup_{1 \leq j \leq 2M(N)} \left| \int_{\gamma_j} \frac{d\phi}{\phi} \right| < \infty.$$

Then for each  $i \in \mathbb{Z}^+$ , we have

$$\sum_{k=1}^{\infty} \int_{Q_k}^{P_k} d\omega_i + K \int_{Q_\infty}^{P_\infty} d\omega_i = 0 \quad \text{mod } \pi \quad (36)$$

Here the paths of integration from  $Q_k$  to  $P_k$  are taken to lie in  $\lambda^{-1}(\bar{I}_k)$ . For the index  $\infty$  the path lies in  $\lambda^{-1}((-\infty, \tau_0))$ .

*Proof.* We note that since  $\phi$  is a meromorphic function on  $\mathcal{R}$  it gives rise to an abelian differential  $d\phi/\phi$  and  $\phi'/\phi$  in local coordinates has poles precisely at the points  $P_i$ ,  $Q_i$ ,  $P_\infty$  and  $Q_\infty$ , hence its divisor is of the form in (35). Since only finitely many of the  $P_i$ ,  $Q_i$  fall in  $\mathcal{R}_N$ ,  $d\phi/\phi$  has finitely many poles in  $\mathcal{R}_N$ .

We pass to the simply connected region  $E_N$  obtained from the normal polygon of  $\mathcal{R}_N$ . We note that in this region the differentials  $d\omega_i$  are exact. Therefore the integrals  $\omega_i(z) = \int_{z_0}^z d\omega_i$  are analytic functions in  $E_N$  as long as the paths of integration lie in the interior of  $E_N$ . Then  $\omega_i d\phi/\phi$  is also a meromorphic differential and has poles precisely at the poles of  $d\phi/\phi$ . We consider the integral of  $\omega_i d\phi/\phi$  along the boundary  $C$  of  $E_N$ , slightly deformed if necessary, such that none of the poles of the differential fall on it. Then we have, by applying the residue theorem, in the simply connected region  $E_N$ ,

$$\begin{aligned} \frac{1}{2\pi i} \int_C \omega_i \frac{d\phi}{\phi} &= \sum_{j=1}^{k(N)} [\omega_i(P_j) - \omega_i(Q_j)] + K[\omega_i(P_\infty) - \omega_i(Q_\infty)] \\ &= \sum_{k=1}^{k(N)} \int_{Q_i}^{P_i} d\omega_i + \int_{Q_\infty}^{P_\infty} d\omega_i \end{aligned} \quad (37)$$

We compute the quantities on the left directly. The integral over the curve  $C$  can be written as

$$\int_C \omega_i \frac{d\phi}{\phi} = \sum_{i=1}^{k(N)} \int_{a_i \cup b_i \cup a_i^{-1} \cup b_i^{-1}} \omega_i \frac{d\phi}{\phi} + \sum_{j=1}^{2k(N)} \int_{\gamma_j \cup \gamma_j^{-1} \cup \partial H_j} \omega_i \frac{d\phi}{\phi} \quad (38)$$

by a slight deformation of the contour if necessary. To compute this integral explicitly we use the following properties of the values of  $\omega_i$  on the corresponding points on the  $a$ ,  $b$ ,  $\gamma$ 's and  $a^{-1}$ ,  $b^{-1}$ ,  $\gamma^{-1}$ 's. We denote the points on  $a_j^{-1}$  corresponding to the

point  $p$  in  $a_j$  as  $p^*$  and similarly for the  $b$ 's and the  $\gamma$ 's.

$$\frac{\phi'}{\phi}(p^*) = \frac{\phi'}{\phi}(p), \quad (39)$$

$$\omega_i(p^*) = \begin{cases} \omega_i(p) + \pi\delta_{ij} & p \in a_j \\ \omega_i(p) + \pi_{ij}, & p \in b_j \\ \omega_i(p) + \int_{\partial H_j} \omega_i, & p \in \gamma_j. \end{cases}$$

Using these relations, and the fact integrating  $(\phi')/\phi$  over an alpha or a beta cycle gives the change in the argument of  $\phi$ , which is  $2\pi i$  times an integer, we compute the integrals in (38) to obtain the value of the integral as

$$\int_{a_i \cup b_i \cup a_i^{-1} \cup b_i^{-1}} \omega_i \frac{d\phi}{\phi} = \sum_{j=1}^{k(N)} \pi\delta_{ij}(2\pi i m_j) + \pi_{ij}(2\pi i l_j) \quad (40)$$

for some integers  $m_j$  and  $l_j$ . Therefore collecting the above equations together we have,

$$\begin{aligned} \sum_{j=1}^{k(N)} [\omega_i(P_j) - \omega_i(Q_j)] &= -K[\omega_i(P_\infty) - \omega_i(Q_\infty)] + \pi\delta_{ij}m_i \\ &+ \sum_{j=1}^{k(N)} \pi_{ij}l_j + e_N \end{aligned} \quad (41)$$

where we have taken

$$e_N = \frac{1}{2\pi i} \left\{ \sum_{j=1}^{2M(N)} \int_{\partial H_j} \omega_i \frac{d\phi}{\phi} + \sum_{j=1}^{2M(N)} \left( \int_{\partial H_j} \omega_i \right) \left( \int_{\gamma_j} \frac{d\phi}{\phi} \right) \right\}.$$

We can compute the left hand side of the above equation on  $\mathcal{R}$  to get

$$\begin{aligned} \sum_{j=1}^{k(N)} \int_{Q_j}^{P_j} d\omega_i &= -K \int_{Q_\infty}^{P_\infty} d\omega_i + \pi\delta_{ij}m_i \\ &+ \sum_{j=1}^{k(N)} \pi_{ij}l_j + e_N. \end{aligned} \quad (42)$$

At this stage of approximation, the sums are finite. We will need only real part of the above equation in the limiting case so that we can take the real parts at this stage. Then the  $\pi_{ij}$  terms drop out, since it is purely imaginary. By assumption, the term  $e_N$  goes to zero and the real part of  $\omega_i(P_j) - \omega_i(Q_j)$  can be evaluated using  $d\omega_i$  with the path of integration in  $\lambda^{-1}(\bar{I}_j)$  and a path that lies in  $\lambda^{-1}((-\infty, \tau_0))$  for the  $j = \infty$ , term, so that the quantities are real.  $\square$

In view of the choice of paths in (36), we use a continuous parametrization for  $\lambda^{-1}(\bar{I}_j)$  and rewrite (36) as follows.

Denote the points  $x_j$  in  $\lambda^{-1}(\bar{I}_j)$  by

$$x_i = \tau_{2i-1} + l_i \sin^2(\theta_i) \quad (43)$$

where

$$0 \leq \theta_i < \frac{\pi}{2}, \quad \text{if } x_i \in \mathcal{R}^+ \quad \text{and} \quad \frac{\pi}{2} \leq \theta_i < \pi, \quad \text{if } x_i \in \mathcal{R}^-.$$

### Inverse spectral theory

In this parametrization we denote the angles corresponding to  $P_j$  and  $Q_j$  as  $\theta_j$  and  $\nu_j$  respectively. Then using the definition of  $d\omega_i$ , in terms of the functions  $f_i$  of the Definition (3.8), equation (34) and change variables to the angular variables  $\theta$ , we have

$$\sum_{k=1}^{\infty} \int_{\psi_k}^{\theta_k} f_{ik}(\theta) d\theta = -Kc_i + \pi m_i. \quad (44)$$

We write the  $f_{ik}$  below for clarity, where we denote  $s(\theta) = \sin^2(\theta)$ .

$$\begin{aligned} f_{ik}(\theta) = & \frac{\sigma_i}{\sqrt{(l_k s(\theta) + \tau_{2k-1} - \tau_0)(l_k s(\theta) + \tau_{2k-1} - \tau_{\infty})}} \\ & \times \frac{(l_s s(\theta) + \tau_{2k-1} - \zeta_k^i)}{\sqrt{(l_k s(\theta) + \tau_{2k-1} - \tau_{2i-1})(l_k s(\theta) + \tau_{2k-1} - \tau_{2i})}} \\ & \times \prod_{j \neq i, k} \frac{(l_k s(\theta) + \tau_{2k-1} - \zeta_j^i)}{\sqrt{(l_k s(\theta) + \tau_{2k-1} - \tau_{2j-1})(l_k s(\theta) + \tau_{2k-1} - \tau_{2j})}}. \end{aligned} \quad (45)$$

We show that  $f_{ik}$  satisfy the following bounds as functions from  $[0, \pi]$  to  $\mathbb{R}$ .

**Lemma 4.4.** *The  $f_{ik}$  defined above are periodic functions of  $\theta$  and they satisfy the following bounds for each  $k \neq i$ .*

$$\sup_{\theta} |f_{ik}(\theta)| \leq C_1 q_k \quad \text{and} \quad \sup_{\theta} |f_{ik} f_{kk}^{-1}(\theta)| \leq C_2 q_k.$$

For  $k = i$ , we have the following upper and lower bounds,

$$C_3 \leq |f_{ii}(\theta)| \leq C_4$$

with the constants  $C_3, C_4$  independent of  $i$ .

*Proof.* The quantities  $f_{ik}$ s are nothing but  $\sqrt{(\lambda - \tau_{2k-1})(\lambda - \tau_{2k})} \sigma_i \omega_{r_i}$  of (32), written in  $\theta$  variables, so that the estimates are clear from the estimates of lemma (3.2) and the estimates for  $\sigma_i$  coming from (32).  $\square$

At this stage we state a corollary of the interpolation theorem that will be useful in the next subsection in showing the invertibility of a linear map.

#### COROLLARY 4.5

Consider a set of points  $\theta_j$  one in each  $[0, \pi/2]$  or  $[\pi/2, \pi]$ ,  $j = 1, 2, \dots$ . Then for each  $j$ , there exists a sequence of numbers  $C_i^j$ , with  $C_i^j \leq C q_i$ ,  $i = 1, 2, \dots$ , such that,  $\sum_{i=1}^{\infty} C_i^j f_{ik}(\theta_k) = 0$  for each  $k \neq j$  and  $\sum_{i=1}^{\infty} C_i^j f_{ij}(\theta_j) \neq 0$ .

*Proof.* Consider the point  $x \in \Psi$  given by,

$$x_i = \tau_{2i-1} + l_i \sin^2(\theta_i)$$

for any given  $j$  fixed. Consider,  $C_i^j$  given by,

(46)

Then using the bounds (5) of lemma (3.2), we have the estimate

$$C_i^j = \int_{I_i} \omega_{x^j}(\lambda) d\lambda \leq C q_i \quad C_1 \leq C_j^j = \int_{I_j} \omega_{x^j}(\lambda) d\lambda \leq C_2.$$

with the constants independent of  $i$ . This implies that for fixed  $j$ , the vector  $C^j$  is in  $\mathcal{H}$ . Therefore by the interpolation theorem, for each  $k$ , we have the relation,

$$\sqrt{(\lambda - \tau_{2k-1})(\lambda - \tau_{2k})} \omega_{x^j}(\lambda) = \sum_{i=1}^{\infty} C_i^j f_i(\lambda) \sqrt{(\lambda - \tau_{2k-1})(\lambda - \tau_{2k})}.$$

Evaluating the left hand side at the points  $x_k^j$  and writing the right hand side of the above equation using equations (45, 46), we have the required statements at the points  $\theta_k$ , since the left hand side vanishes at the points  $x_k^j$  for each  $k \neq j$ . For  $k = j$ , the left hand side is non-zero. Clearly this argument is valid for any  $j$ .  $\square$

#### 4.2 The Abel-Jacobi map

We consider here the Abel-Jacobi map of equation (44) and prove its properties. To start with we consider the following real Banach space. We consider the numbers  $s_i$  of assumption (2.1), consider the finite measure  $\mu = \sum_{i=1}^{\infty} s_i \delta_i$  and consider

$$X = l_{\mathbb{R}}^2(\mathbb{Z}^+, \mu) \quad \text{and} \quad X_{\mathbb{Z}} = \{x \in X : x_i \in \pi \mathbb{Z} \forall i\}. \quad (47)$$

We note that the topology on  $X$  is such that the  $l^{\infty}$  unit ball is compact in  $X$ . Therefore the set  $\mathcal{B}_X = X/X_{\mathbb{Z}}$  is compact in  $X$ .

We define the Abel-Jacobi map  $\mathcal{A}$  from  $X$  to itself as

$$\mathcal{A}_i(\theta) = \sum_{j=1}^{\infty} \int_0^{\theta_j} f_{ij}(\theta) d\theta$$

and split it into two parts, the "diagonal"  $\mathcal{A}_1$  and the "off diagonal"  $\mathcal{A}_2$ , as

$$\mathcal{A}_{1i}(\theta) = \int_0^{\theta_i} f_{ii}(\theta) d\theta, \quad \mathcal{A}_{2i}(\theta) = \sum_{j \neq i}^{\infty} \int_0^{\theta_j} f_{ij}(\theta) d\theta.$$

Henceforth we use a **vector notation** for points of  $X$ ,  $X_{\mathbb{Z}}$  etc., Then  $\mathcal{A}$ ,  $\mathcal{A}_1$  and  $\mathcal{A}_2$  have the following properties.

#### PROPOSITION 4.6

The maps  $\mathcal{A}$ ,  $\mathcal{A}_1$  and  $\mathcal{A}_2$  map  $X$  to itself and satisfy

1.  $\mathcal{A}(\theta + \mathbf{m}) = \mathcal{A}(\theta) + \mathbf{m}$ ,
2.  $\mathcal{A}_1(\theta + \mathbf{m}) = \mathcal{A}_1(\theta) + \mathbf{m}$ ,
3.  $\mathcal{A}_2(\theta + \mathbf{m}) = \mathcal{A}_2(\theta)$ ,

for each  $\theta$  in  $X$  and  $\mathbf{m} \in X_{\mathbb{Z}}$ .

*Proof.* We note that the functions  $f_{ik}(\theta)$  are periodic of period  $\pi$ . Therefore the



normalizations of equation (34), imply that for  $k \neq i$ ,

$$\int_0^{\theta+\pi l} f_{ik}(\theta) = \int_0^{\theta} f_{ik}(\theta) \quad \forall l \in \mathbb{Z}$$

and

$$\int_0^{\theta+\pi l} f_{ii}(\theta) = \int_0^{\theta} f_{ii}(\theta) + \pi l \quad \forall l \in \mathbb{Z}.$$

These periodicity relations together with the uniform boundedness of the functions  $f_{ik}$  when  $\theta$  varies over  $[0, \pi]$ , shows that  $\mathcal{A}$ ,  $\mathcal{A}_1$  and  $\mathcal{A}_2$  map  $X$  to itself. The stated periodicity relations of the proposition are again easy consequences of the above periodicity of  $f_{ik}$ s and the normalisations of proposition (3.8).  $\square$

We note that the periodicity of  $\mathcal{A}_2$  shows that it is a bounded map from  $X$  to itself. This will be useful in the following.

We would like to show that the map  $\mathcal{A}$  is a Lipschitz continuous bijection from  $X$  to itself, with its inverse also being Lipschitz. To do this we consider an auxiliary map

$$\mathcal{B} = (I + \mathcal{A}_2 \circ \mathcal{A}_1^{-1}).$$

We will show below that  $\mathcal{B}$  is a differentiable, invertible map of  $X$  to itself and its derivative is bounded on  $X$ . Similar properties are true for its inverse. To do this we first consider the matrices  $A_1$ ,  $A_2$  and  $B$  of partial derivatives of  $\mathcal{A}_1$ ,  $\mathcal{A}_2$  and  $\mathcal{A}_2 \circ \mathcal{A}_1^{-1}$ . For a point  $\theta \in X$ .

$$A_1(\theta)_{ij} = f_{ii}(\theta_i) \delta_{ij} \quad A_2(\theta)_{ij} = f_{ij}(\theta_j)(1 - \delta_{ij}) \quad B(\theta) = A_2 \circ A_1^{-1}(\theta). \quad (48)$$

Let  $\mathcal{L}_X$  denote the space of bounded linear operators on  $X$ . In the proposition below the continuity of  $A_1$  from  $X$  to  $\mathcal{L}_X$  is not clear since the equicontinuity of the family  $\{f_{ii}\}$  of maps is not clear.

**Lemma 4.7.** *Consider  $X$  and  $\mathcal{L}_X$  as metric spaces equipped with the respective norm topologies. Then the maps  $A_1(\theta)$  are bounded from  $X$  to  $\mathcal{L}_X$ . The maps  $A_2(\theta)$  and  $B(\theta)$  are compact operator valued and are continuous from  $X$  to  $\mathcal{L}_X$ .*

*Proof.* The boundedness of  $A_1(\theta)$  as a linear operator from  $X$  to  $X$ , for any  $\theta$  is clear from the estimates of lemma (4.4). The estimates of lemma (4.4), and the uniform boundedness of  $q_i/s_i$  coming from assumption (2.2) show that for any  $\psi \in X$ ,

$$|(A_2(\theta)\psi)_i| \leq \sum_{k=1}^{\infty} |f_{ik}(\theta_k)\psi_k| \leq C \sum_{k=1}^{\infty} s_k |\psi_k| = C \|\psi\|.$$

This estimate shows that  $A_2(\theta)$  maps a bounded subset of  $X$  into a precompact subset of  $X$  for each fixed  $\theta$ , showing the compactness of  $A_2(\theta)$ . The estimate also shows that the norms  $\|A_2(\theta)\|$  are uniformly bounded on  $X$ . The proof for  $B(\theta)$  follows from compactness of  $A_2$  and the boundedness of  $A_1^{-1}$  coming from the bounds of lemma (4.4). We will show the continuity of  $A_2(\theta)$ , the proof for  $B$  is similar using the estimates of lemma (4.4). Consider and  $\psi$  in  $X$  and consider  $\theta, \phi \in X$  and an  $\varepsilon$

fixed. Then we have

$$\begin{aligned}
& \| (A_2(\theta) - A_2(\phi)) \psi \| \\
&= \sum_{i=1}^{\infty} s_i \left\| \sum_{k \neq i}^{\infty} (f_{ik}(\theta_k) - f_{ik}(\phi_k)) \psi_k \right\| \\
&= \sum_{i=1}^{M(\varepsilon)} s_i \left\| \sum_{k(\neq i)=1}^{l(\varepsilon)} (f_{ik}(\theta_k) - f_{ik}(\phi_k)) \psi_k \right\| \\
&\quad + \sum_{i=1}^{M(\varepsilon)} s_i \left\| \sum_{k(\neq i)=l(\varepsilon)+1}^{\infty} (f_{ik}(\theta_k) - f_{ik}(\phi_k)) \psi_k \right\| \\
&\quad + \sum_{i=M(\varepsilon)+1}^{\infty} s_i \left\| \sum_{k \neq i}^{\infty} (f_{ik}(\theta_k) - f_{ik}(\phi_k)) \psi_k \right\|. \tag{49}
\end{aligned}$$

Given  $\varepsilon$  we can choose  $M(\varepsilon)$  so that the last two terms in the above inequality are bounded by  $C\varepsilon\|\psi\|$ , using the summability of  $s_i$ , the bounds of lemma (4.4) for  $f_{ik}s$  and the estimate  $q_k \leq C s_k$  coming from the summability of  $q_k/s_k$ . Therefore we concentrate on the first term in the inequalities. We can find, for each  $i = 1, \dots, M(\varepsilon)$  and  $k = 1, \dots, l(\varepsilon)$ , numbers  $\delta_{ik}$  so that, by continuity of  $f_{ik}s$ ,

$$|f_{ik}(\theta_k) - f_{ik}(\phi_k)| \leq s_k \varepsilon \quad \text{whenever} \quad |\theta_k - \phi_k| < \delta_{ik}.$$

Now choose  $\delta$  so that

$$\delta = \inf_{\substack{i=1, \dots, M(\varepsilon) \\ k(\neq i)=1, \dots, l(\varepsilon)}} \{ \delta_{ik}, s_k \delta_{ik} \}.$$

Then it follows that if  $\|\theta - \phi\| < \delta$ , we have for each  $i = 1, \dots, M(\varepsilon)$ ,  $k \neq i = 1, \dots, l(\varepsilon)$ ,

$$s_k |\theta_k - \phi_k| < \delta \leq s_k \delta_{ik} \Rightarrow |\theta_k - \phi_k| < \delta_{ik} \Rightarrow |f_{ik}(\theta_k) - f_{ik}(\phi_k)| \leq s_k \varepsilon.$$

These estimates show that the first term in the inequalities (49) are bounded by  $C\varepsilon\|\psi\|$  again by the summability of  $s_i s$ . These inequalities together show that given  $\varepsilon$  positive we have a  $\delta$  such that,

$$\| (A_2(\theta) - A_2(\phi)) \psi \| \leq C\varepsilon\|\psi\|$$

whenever  $\|\theta - \phi\| < \delta$ . □

**Theorem 4.8.** Consider the maps  $\mathcal{A}, \mathcal{A}_1$  and  $\mathcal{B}$ . Then we have the following properties.

1.  $\mathcal{A}_1, \mathcal{A}_1^{-1}$  are Lipschitz continuous bijections of  $X$ .
2.  $\mathcal{B}$  is differentiable from  $X$  to itself with the derivative invertible and uniformly bounded on  $X$ .
3.  $\mathcal{B}$  and  $\mathcal{A}$  are Lipschitz continuous bijections of  $X$ .
4.  $\mathcal{A}^{-1}$  is a Lipschitz continuous map of  $X$  to itself.

*Proof.*

1. The property (1) of proposition (4.6), shows that it is enough to consider the map  $\mathcal{A}_1$  on  $\mathbb{B}_X$ . For each  $i$ , the continuous map  $g_i(\theta_i) = (\mathcal{A}_1(\theta))_i$ , satisfies,  $g_i(0) = 0$ ,

$g_i(\pi) = \pi$  is strictly monotone on  $[0, \pi]$ , hence it is a bijection of  $[0, \pi]$  to itself. This shows that  $\mathcal{A}_1$  is a bijection of  $\mathbb{B}_X$  to itself. On the other hand, the functions  $f_{ii}(\theta_i)$  and  $f_{ii}^{-1}(\theta_i)$  are uniformly bounded above and below, by lemma (4.4), showing that both  $\mathcal{A}_1$  and  $\mathcal{A}_1^{-1}$  are Lipschitz on  $\mathbb{B}_X$ .

2. We prove here that  $\mathcal{A}_2$  is differentiable, the proof for  $\mathcal{A}_2 \circ \mathcal{A}_1^{-1}$  is similar. Let  $A_2(\theta)$  denote the matrix of partial derivatives of  $\mathcal{A}_2$  at the point  $\theta$  in  $X$ . Then we have for any  $h \in X$  (indeed we can take  $h$  in  $\mathcal{B}_X$  without loss of generality by the periodicity of  $\mathcal{A}_2$ ),

$$\begin{aligned} & \|\mathcal{A}_2(\theta + h) - \mathcal{A}_2(\theta) - A_2(\theta)h\| \\ &= \sum_{i=1}^{\infty} s_i |\mathcal{A}_2(\theta + h)(i) - \mathcal{A}_2(\theta)(i) - (A_2(\theta)h)(i)| \\ &= \sum_{i=1}^{\infty} s_i \left| \sum_{k \neq i} \left[ \int_{\theta_k}^{\theta_k + h_k} f_{ik}(\phi_k) d\phi_k - f_{ik}(\theta_k) h_k \right] \right| \\ &= \sum_{i=1}^{\infty} s_i \left| \sum_{k \neq i} [f_{ik}(\tilde{\theta}_k) - f_{ik}(\theta_k)] h_k \right| \\ &\leq \|A_2(\tilde{\theta}) - A_2(\theta)\| \|h\|. \end{aligned} \tag{50}$$

We have used the mean value theorem for the functions  $\int f_{ik}$  in the second step above and choose the point  $\tilde{\theta}_k$  from  $[\theta_k, \theta_k + h_k]$ . Now the continuity of  $A_2$ , mentioned in lemma shows that as  $\|h\|$  goes to zero,  $\tilde{\theta} \rightarrow \theta$  in  $X$ , showing the differentiability, by lemma (4.7). Similarly the differentiability of  $\mathcal{A}_2 \circ \mathcal{A}_1^{-1}$  and hence that of  $\mathcal{B}$  is shown. Therefore, by lemma (4.7) the derivative  $I + B(\theta)$  of  $\mathcal{B}$  at  $\theta$  is uniformly bounded and continuous from  $X$  to  $\mathcal{L}_X$ . To show the invertibility of  $(I + B(\theta))$ , it is enough to show that zero is not an eigenvalue, by the Fredholm alternative, since  $B$  is a compact operator. Suppose there is a  $\phi$  in  $X$  such that

$$(I + B(\theta))\phi = 0.$$

Then, since  $A_1$  is an invertible operator, it follows that,

$$(A_1(\theta) + A_2(\theta))\psi = 0$$

for some  $\psi$  in  $X$ . Writing the above equation explicitly in terms of the matrix elements, we have that

$$\sum_{j=1}^{\infty} f_{ij}(\theta_j) \psi_j = 0 \quad \forall i.$$

Then by corollary (4.5), we can choose numbers  $C_i^k$  such that after an interchange of sums we get,

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} C_i^k f_{ij}(\theta_j) = \left( \sum_{i=1}^{\infty} C_i^k f_{ik}(\theta_k) \right) \psi_k = 0 \quad \forall k.$$

Since the sum in the parantheses is non-zero by lemma (4.5), we have that  $\psi_k = 0$  for each  $k$ . The continuity of  $(I + B(\theta))$  as a map from  $X$  to  $\mathcal{L}_X$  implies, by an

application of the resolvent equation, the inverse is also continuous and hence uniformly bounded on compacts of  $X$ . Further, the periodicity of the operator valued function  $B(\theta)$ , with its periods in  $X_{\mathbb{Z}}$ , coming from the periodicity of  $f_{ij}$ s, shows that  $\|(I + B(\theta))^{-1}\|$  is uniformly bounded on  $X$ .

3. To show that  $\mathcal{B}$  and  $\mathcal{A}$  are bijections of  $X$  to itself, we note that by (1), it is enough to show that  $\mathcal{A}$  is injective and  $\mathcal{B}$  is surjective. Suppose there are two points  $\theta$  and  $\psi$  in  $X$  such that  $\mathcal{A}(\theta) = \mathcal{A}(\psi)$ . Let  $\eta$  and  $\zeta$  be in  $\Psi$  such that  $\eta_j = \tau_{2j-1} + l_j \sin^2(\theta_j)$  and  $\zeta_j = \tau_{2j-1} + l_j \sin^2(\psi_j)$ . Then  $\mathcal{A}(\theta) = \mathcal{A}(\psi)$  implies,

$$\sum_{j=1}^{\infty} \int_{\eta_j}^{\zeta_j} f_i(\lambda) d\lambda = 0.$$

for each  $i$ . The interpolation theorem implies that

$$\sum_{j=1}^{\infty} \int_{\eta_j}^{\zeta_j} f(\lambda) d\lambda = 0.$$

for each  $f \in \mathcal{H}_{\infty-1}$ . Then corollary (3.10) implies that  $\eta_j = \zeta_j$  for each  $j$ . This implies in turn that  $\theta$  and  $\psi$  differ at most by an element of  $X_{\mathbb{Z}}$  so that  $\theta = \psi + \mathbf{m}$ . Using proposition (4.6(1)) we see that,

$$\mathcal{A}(\theta) = \mathcal{A}(\psi + \mathbf{m}) = \mathcal{A}(\psi) + \mathbf{m}$$

which implies that  $\mathbf{m}$  is zero.

The differentiability of  $\mathcal{B}$  allows us to use the inverse function theorem on  $X$ , to show that the range of  $\mathcal{B}$  is open. We will show that the range of  $\mathcal{B}$  is closed showing, by connectedness of  $X$ , that  $\mathcal{B}$  is onto. The map  $\mathcal{B}$  satisfies  $\mathcal{B}(\theta + \mathbf{m}) = \mathcal{B}(\theta) + \mathbf{m}$  for all  $\mathbf{m} \in X_{\mathbb{Z}}$ . We also have that if  $x \in \text{Ran } \mathcal{B}$  and  $\mathbf{m} \in X_{\mathbb{Z}}$ , then  $x + \mathbf{m} \in \text{Ran } \mathcal{B}$ . Consider  $y^n$  in  $\text{Ran } \mathcal{B}$ , converging to some  $y$  in  $X$ . Then, decompose  $y^n$  and  $y$  as  $(y^n) + [y^n]$ ,  $(y) + [y]$  into their parts in  $\mathbb{B}_X$  and  $X_{\mathbb{Z}}$ . Notice that  $[y^n] - [y] = y^n - y - (y^n) + (y)$ . Since  $(y^n) - (y)$  belongs to a compact set, some subsequence of this converges, say to  $x$  and it is clear that  $x \in X_{\mathbb{Z}}$ . It is also clear that  $[y^n] - [y] - x$  goes to zero for this subsequence. Therefore  $y$  is in  $\text{Ran } \mathcal{B}$  follows from showing that  $\text{Ran}(\mathcal{B}) \cap \mathbb{B}_X$  is closed. Let  $\mathcal{C}$  denote  $\mathcal{A}_2 \circ \mathcal{A}_1^{-1}$ , then by lemma (4.6),  $\mathcal{C}$  is also a periodic bounded map of  $X$  to itself with the coordinate maps  $\mathcal{C}_i$  uniformly bounded on  $X$ . Consider a sequence  $\theta^n \in \text{Ran}(\mathcal{B}) \cap \mathbb{B}_X$  converging to  $\theta$  in  $\mathbb{B}_X$ . Then the sequence  $\psi^n$  of vectors in  $X$ , with  $\mathcal{B}(\psi^n) = \theta^n$ , are in a compact set, since,

$$\theta_i^n = \mathcal{B}(\psi^n)_i = \psi_i^n + \mathcal{C}(\psi^n)_i \Rightarrow \psi_i^n = \theta_i^n - \mathcal{C}(\psi^n)_i$$

and  $\mathcal{C}(\psi^n)_i$  are bounded uniformly the bound independent of  $\psi^n$  and  $i$ . Hence there is a subsequence  $\psi^{n_k}$  of  $\psi^n$  that converges to some point  $\psi$  in  $X$ . Therefore by the continuity of  $\mathcal{B}$ , it is clear that  $\theta$  is precisely the image of  $\psi$  under  $\mathcal{B}$ .

4. The inverse  $\mathcal{B}^{-1}$  of  $\mathcal{B}$  exists by (4) and it is also differentiable with the derivative  $(I + B(\theta))^{-1}$  uniformly bounded on  $X$  by (2) so it is Lipschitz. The Lipschitz continuity of  $\mathcal{A}$  now follows from those of  $\mathcal{B}$  and  $\mathcal{A}_1$ . □

Finally we end this section showing the almost periodicity of the preimages of the orbits of  $nc$ , for a given vector  $c$  in  $X$ .

**orem 4.9.** Consider any two vectors  $\mathbf{c}$  and  $\mathbf{d}$  in  $X$ . Consider the sequence of points  $n) \in \Psi$  with  $\xi_i(n) = \tau_{2i-1} + l_i \sin^2(\mathcal{A}_i^{-1}(nc + \mathbf{d}))$ ,  $n \in \mathbb{Z}$ . Then  $\xi_i(n)$  is an almost periodic quence, in  $n$ , for each  $i$ .

*oof.* Let  $y_i(n) = nc_i + d_i \bmod \pi$ , then  $\sin^2(\mathcal{A}_i^{-1}(nc + \mathbf{d})) = \sin^2(\mathcal{A}_i^{-1}(y(n)))$ . Therefore the stated almost periodicity, it is enough to consider  $y(n)$ . Given an  $\varepsilon$  we choose  $K(\varepsilon)$  and truncate  $y(n)$  to  $y^K(n)$  where  $y^K(n)_i$  is zero for  $i > K$  and equals  $y_i(n)$  for  $i \leq K$ , so that  $\|y(n) - y^K(n)\| < \varepsilon$ . Then  $y^K(n)$  is an element of  $\mathbb{R}^K / \pi \mathbb{Z}^K$ , on which it satisfies  $\sup_n \|y^K(n) - y^K(n + N)\| < \varepsilon$  for an  $N$  depending upon  $\varepsilon$  and  $K(\varepsilon)$ , by an application of Poincare recurrence theorem. Therefore by the Lipschitz continuity of  $\sin^2 \circ \mathcal{A}_i$ , coming from the Lipschitz continuity of  $\sin^2$  and  $\mathcal{A}_i^{-1}$  we have that

$$\begin{aligned} \sup_n |\xi_i(n) - \xi_i(n + N)| &\leq \frac{l_i}{s_i} |s_i| |\mathcal{A}_i^{-1}(y(n)) - \mathcal{A}_i^{-1}(y(n + N))| \\ &\leq q_i \|\mathcal{A}_i^{-1}(y(n)) - \mathcal{A}_i^{-1}(y(n + N))\| \\ &\leq C q_i \|y(n) - y(n + N)\| \\ &< C q_i \|y^K(n) - y^K(n + N)\| + \varepsilon < D\varepsilon \end{aligned}$$

which is the almost periodicity claimed in the theorem.  $\square$

## Random Jacobi matrices

Consider  $(\Omega, \mathcal{B}, \mathbb{P})$ , where  $\Omega =$  the space of  $\mathbb{R}^+ \times \mathbb{R}$  valued bounded sequences  $(a_n, b_n)$ ,  $\mathcal{B} =$  the Borel  $\sigma$ -algebra on  $\Omega$  generated by the topology of point wise convergence and  $\mathbb{P}$  a probability measure on  $(\Omega, \mathcal{B})$ . Let  $T$  denote the translation  $T$  on  $\Omega$  given by  $(T(a, b))(n) = (a_{n-1}, b_{n-1})$ . Then corresponding to each  $\omega \in \Omega$  we have a self-adjoint operator acting on  $l^2(\mathbb{Z})$  given by, [5]

$$(H^\omega u)(n) = a_{n-1}^\omega u(n-1) + a_n^\omega u(n+1) + b_n^\omega u(n) \quad u \in l^2(\mathbb{Z}) \quad (51)$$

It was proved by Pastur [33] under the assumptions on  $\mathbb{P}$ , that the spectrum  $\Sigma$  of the operators  $H^\omega$  is a non random set. It was shown by Kunz-Souillard [25] that the spectral type of the operators is also non random, that is there are non random sets  $\Sigma_{ac}$  and  $\Sigma_{sc}$  and  $\Sigma_{pp}$  which are respectively the absolutely continuous, singular continuous and pure point spectra of the operators  $H^\omega$  for almost all  $\omega$ .

Consider the operator  $H^\omega$  for  $\omega$  in support of  $\mathbb{P}$ . The following facts are contained essentially in Carmona–Kotani [5] and in the proofs of Simon [35]. These are in fact worked out in the book of Carmona–Lacroix [6].  $H^\omega$  is a bounded self-adjoint operator and hence  $\lambda \in \mathbb{C}^+ \cup \mathbb{C}^-$  belongs to the resolvent of  $H^\omega$ . The Wronskian of two solutions  $f, g$  of the eigenvalue equation  $(H^\omega - \lambda)u = 0$  is given by

$$W(f, g) = a_n [f(n+1)g(n) - f(n)g(n+1)]$$

and is a constant. By Weyl theory, there are unique solutions, for the eigenvalue equation  $(H^\omega - \lambda)u = 0$  which are integrable at  $\pm \infty$  for  $\lambda$  in the resolvent set of  $H^\omega$ . We denote them by  $u_{\pm, \lambda}^\omega$ . Then we have the expression for Green function given by

$$g_\lambda^\omega(n, n) = \frac{u_{+, \lambda}^\omega(n) u_{-, \lambda}^\omega(n)}{W^\omega} \quad (52)$$

where  $W^\omega$  is the Wronskian of the solutions  $u_{\pm,\lambda}^\omega$  given by  $a_0^\omega(u_{+,\lambda}^\omega(1)u_{-,\lambda}^\omega(0) - u_{+,\lambda}^\omega(0)u_{-,\lambda}^\omega(1))$ . The Weyl functions given by

$$m_+^\omega(\lambda)(n) = -\frac{u_{+,\lambda}^\omega(n+1)}{a_n^\omega u_{+,\lambda}^\omega(n)} \quad \text{and} \quad m_-^\omega(\lambda)(n) = -\frac{u_{-,\lambda}^\omega(n-1)}{a_{n-1}^\omega u_{-,\lambda}^\omega(n)} \quad (53)$$

are related to the operators  $H_{\pm,n}^\omega$  of  $H^\omega$  restricted to the subspaces  $l^2[n+1, \infty)$  and  $l^2(-\infty, n-1]$  via

$$m_{\pm,n}^\omega(\lambda) = (H_{\pm,n}^\omega - \lambda)^{-1}(n \pm 1, n \pm 1).$$

We define the scaled  $m$ -functions  $M_\pm$  by

$$M_{+,n}^\omega(\lambda) = a_n^2 m_{+,n}^\omega(\lambda) \quad \text{and} \quad M_{-,n}^\omega(\lambda) = a_{n-1}^2 m_{-,n}^\omega(\lambda). \quad (54)$$

Then the  $M_\pm$  satisfy the following equations

$$\begin{aligned} M_{+,n}^\omega(\lambda) &= b_n^\omega - \lambda - (a_{n-1}^\omega)^2 (M_{+,n-1}^\omega)^{-1} \quad \text{and} \\ M_{-,n}^\omega(\lambda) &= b_n^\omega - \lambda - (a_n^\omega)^2 (M_{-,n+1}^\omega)^{-1} \end{aligned} \quad (55)$$

The  $M$ -functions and the Green function are related by,

$$g_\lambda^\omega(n, n) = \frac{-1}{M_{+,n}^\omega(\lambda) + M_{-,n}^\omega(\lambda) + \lambda - b_n^\omega}. \quad (56)$$

Under the Assumptions on the probability measure  $\mathbb{P}$ , the Lyapunov exponent  $\gamma(\lambda)$  exists for all  $\lambda \in \mathbb{C}^+$  and is related to the  $m$ -functions by,

$$\mathbb{E}_{\mathbb{P}} \{\log |M_{+,0}(\lambda)|\} = -\gamma(\lambda) + \mathbb{E}_{\mathbb{P}} \{\log a_0\}. \quad (57)$$

Then by a combination of theorems of Ishii-Pastur, Kotani and Simon it follows that

$$\Sigma_{ac}(H^\omega) = \{E : \gamma(E + i0) = 0\}^{-\text{ess}}$$

— ess denoting closure up to sets of Lebesgue measure zero. There is also a probability measure on  $\mathbb{R}$  supported on the spectrum called the density of states  $dn$  and the following Thouless formula relating the Weyl functions and the density of states is valid ([5] and [7]).

$$\mathbb{E}_{\mathbb{P}} \{\log M_{+,0}(\lambda)\} = \frac{1}{\pi} \int_{\mathbb{R}} \log \frac{1}{\xi - \lambda} dn(\xi). \quad (58)$$

For a sequence  $\omega_m$  converging to  $\omega$  in the topology of  $\Omega$ , the operators  $H^{\omega_m}$  converge strongly, since  $H^\omega$  are bounded operators with a uniform bound on their norms, by assumption, for  $\omega \in \Omega$ . Therefore the Green functions converge compact uniformly in  $\mathbb{C}^+$ . Using this fact one can get the following theorem of Kotani, on the lines of Kotani [18] Simon [35] or Craig [8].

**Theorem 5.1.** *Let  $\mathbb{P}$  satisfy assumption (3) and let the spectrum of  $H^\omega$  satisfy assumptions (1, 2). Then everywhere on the interior of the spectrum  $\Sigma$  of  $H^\omega$ , the following relation*

s valid

$$\operatorname{Im}[M_{+,n} - M_{-,n}](\lambda + i0) = 0, \quad \operatorname{Re}[M_{+,n}(\lambda + i0) + M_{-,n}(\lambda + i0) + \lambda - b_n] = 0 \quad (59)$$

for all  $\omega$  in support of  $\mathbb{P}$ .

We consider next the trace formulae, well known in the periodic examples and recently constructed for the general bounded Jacobi matrices by Gesztesy–Holden–Simon–Zhao [12]. In the following we set

$$g_\lambda(n, n) \equiv (H - \lambda)^{-1}(n, n), \quad n \in \mathbb{Z}.$$

We however state the trace formula for a special case we are interested in.

**Theorem 5.2.** (trace formula). *Consider a Jacobi matrix given in equation (1) with its spectrum  $\Sigma$  satisfying the assumptions (1). Suppose the Green function  $g_\lambda(n, n)$  has vanishing real part almost everywhere on  $\Sigma$ . Then there is a unique point  $\xi(n) \in \Psi$  such that*

$$b_n = \frac{1}{2}(\tau_0 + \tau_\infty) + \frac{1}{2} \sum_{i=1}^{\infty} (\tau_{2i-1} + \tau_{2i} - 2\xi_i(n))$$

$$a_n^2 + a_{n-1}^2 = \frac{1}{2}(b_n)^2 + \frac{1}{4}(\tau_0^2 + \tau_\infty^2) + \frac{1}{4} \sum_{i=1}^{\infty} \tau_{2i-1}^2 + \tau_{2i}^2 - 2\xi_i(n)^2 \quad (60)$$

*Proof.* We note that the Green function is real and increasing in the gaps  $I_i$  so that it has at most one zero in each gap. When there is a zero in  $I_i$  we call the zero to be  $\xi_i(n)$ , otherwise we take  $\xi_i(n)$  to be  $\tau_{2i-1}$  or  $\tau_{2i}$  according as  $g_\lambda(n, n)$  is positive or negative in  $I_i$ . With this choice we have a point  $\xi(n) \in \Psi$  and we also have by assumption that the Green function has vanishing real part a.e. on the spectrum  $\Sigma$ . Therefore  $g_{\xi(n)}$  constructed, from  $\Sigma$  and  $\xi(n)$ , in proposition (2.2) agrees with  $g_\lambda(n, n)$  everywhere in  $\Pi$ . Therefore the product representation of (12) is valid for  $g_\lambda(n, n)$  in  $\Pi$ . The coefficients of  $1/\lambda^2$  and  $1/\lambda^3$  in the asymptotic expansion of  $g_\lambda(n, n)$  are respectively the left hand sides of the first two relations of the Trace formula stated in the theorem. The right hand sides are the respective coefficients coming from the product representation.  $\square$

### 5.1 Ergodic potentials with band spectrum

In the section we show the existence of probability measures satisfying our Assumptions (3) such that their spectra satisfy the Assumptions (1, 2). We use the inverse spectral theory of Carmona–Kotani [5] to explicitly construct such measures given the spectrum and also present the theorem of Kotani on the ergodic selection of such a measure. Since the proofs of these theorems are essentially contained in the works cited above we do not give proof for theorem (5.4). This theory appears also in the book of Carmona–Lacroix [6].

**Theorem 5.3.** *Consider the set  $\Sigma$  satisfying Assumptions (2). Then there exists a Haralutz*

1.  $w$  and  $w'$  are Herglotz with  $w$  having the representation,

$$w(\lambda) = \frac{1}{\pi} \int \log \frac{1}{\xi - \lambda} dn(\xi)$$

for an absolutely continuous measure  $dn$  supported on  $\Sigma$ .

2.  $w(\lambda) \sim \log -1/\lambda$  as  $\lambda \rightarrow \infty$

3.  $w(\mathbb{C}^+) \subseteq (-\infty, c] \times i[0, \pi]$ , and on  $\Sigma$ ,  $-\gamma = \operatorname{Re}(w) - c = 0$  for some finite  $c$  and  $\int_{\mathbb{R}} \gamma(\xi) dn(\xi) = 0$ .

*Proof.* Consider the set  $\Sigma$  as in Assumption 2.1 and consider the point  $\zeta \in \Psi$  chosen in proposition (3.9) and the corresponding  $h_{\zeta}$  of proposition (3.9). Then, the estimates of proposition (5.6), which are also valid for  $h_{\zeta}$  will show that the measure  $dn$  in the representation

$$h_{\zeta}(\lambda) = \frac{1}{\pi} \int k(\xi, \lambda) dn(\xi)$$

is absolutely continuous, supported on  $\Sigma$  and the distribution function  $n(\xi) = dn(-\infty, \xi]$  satisfies  $n(\infty) = \pi$ . Therefore  $dn(\xi) = \operatorname{Im} h_{\zeta}(\xi + i0) d\xi$ . Using  $dn$  we consider another Herglotz function  $w$  such that

$$w(\lambda) = \frac{1}{\pi} \int_{\mathbb{R}} \log \frac{1}{(\xi - \lambda)} dn(\xi) \quad (61)$$

and show that this is the function stated in the theorem. The above function is well defined for  $\lambda$  in  $\mathbb{C}^+$ . A computation shows that  $w'(\lambda) = h_{\zeta}(\lambda)$  for  $\lambda \in \mathbb{C}^+$ . Since  $\operatorname{Im} h_{\zeta}(\lambda + i0)$  has at most inverse square root singularity at the boundary points of  $\Sigma$ , as can be seen from its product representation,  $\log |\xi - \operatorname{Re} \lambda|$  is integrable with respect to  $dn$ . Therefore by Lebesgue dominated convergence theorem the limits  $w(\xi + i0)$  exist for all  $\xi \in \mathbb{R}$ . We consider the function  $w(\lambda) - w(\tau_0)$  for  $\lambda \in \mathbb{C}^+$ , then the integral representation,

$$g(\lambda) = w(\lambda) - w(\tau_0) = \int_{\tau_0}^{\lambda} w'(\xi) d\xi = \int_{\tau_0}^{\lambda} h_{\zeta}(\xi) d\xi \quad (62)$$

is valid for  $\lambda \in \mathbb{R}$  with the path of integration in  $\mathbb{C}^+$ . The limits  $h_{\zeta}(\lambda + i0)$  exist a.e. so we can take the path of integration along the real axis. Consider the real part of  $g$ . For  $\lambda \in I(-\infty)$  it is obvious that real part of  $g$  is negative. As for  $\lambda$  in a given gap  $I_i$ , we have by proposition (3.9), together with the fact that the real part of  $h_{\zeta}$  in the spectrum is zero, that

$$\operatorname{Re} g(\lambda) = \int_{\tau_{2i-1}}^{\lambda} h_{\zeta}(x) dx$$

is negative, since  $h_{\zeta}$  is an increasing function having at most one zero in the each gap  $I_i$ . Finally the real part of  $g$  in  $I(\infty)$  is negative, since  $h_{\zeta}$  is negative there and the value of  $\operatorname{Re} g$  at any point  $\lambda$  in  $I(\infty)$  is given by  $\int_{\tau_{\infty}}^{\lambda} h_{\zeta}$ , by using proposition (3.9). The analysis to show that the imaginary part of  $g(\lambda)$  is increasing from 0 to  $\pi$  is



ilar. Once we obtain the values of  $w(\lambda)$  on the boundary, by (61), (62) and by the fact that  $\operatorname{Im} w'(\lambda) > 0$  for  $\lambda \in \mathbb{C}^+$  it is clear that

$$\operatorname{Re} w \in (-\infty, c], \quad \operatorname{Im} w \in [0, \pi], \quad c = \omega(\tau_0), \quad \forall \lambda \in \bar{\mathbb{C}}^+$$

the statements of (3) are now clear.  $\square$

Given the  $w$  function satisfying the conditions of Theorem 5.3, we can find a probability measure  $\mathbb{P}$  on  $(\mathbb{R}^+ \times \mathbb{R})^{\mathbb{Z}}$  which is invariant under translations, by the theorem 4.8 of [5]. This shows from their proofs that  $\mathbb{E}_{\mathbb{P}}\{\log a_0\} = c < \infty$ . The periodicity of such a measure follows if the density of states and the Lyapunov exponent satisfy the condition

$$\int \gamma(\xi) dn(\xi) = 0$$

which is true in our case. The proof that this implies the existence of an ergodic measure is almost identical to the proof of Kotani [18], Lemma 7.11 and Theorem 6.3 which can be proved for the  $w$  function satisfying the properties of Theorem 5.3. Since the measure  $dn$  above has compact support  $(\Sigma)$ , the sequences  $(a_n, b_n)$  in support of the measure so constructed will be bounded. Hence we have,

**Theorem 5.4.** *There exists an invariant and ergodic probability measure  $\mathbb{P}$  on  $\Omega$  satisfying assumption (3) having spectrum as in assumptions (1, 2).*

### Almost periodicity

We prove the almost periodicity of the random Jacobi matrices, in this section. We start with getting expressions for the  $M$ -functions and show a technical lemma on their boundedness which verifies some conditions required for the application of Proposition (4.3). We recall the rectangles  $R_N^i$  defined in lemma (3.3) and  $r_N$  defined earlier, we set  $R_N = \bigcup_{i=1}^{M(N)} R_N^i$  for each  $N$ . We consider  $M_{\pm}$  given in (54) and drop the superscript  $\omega$ .

### PROPOSITION 5.5

Consider  $\mathbb{P}$  satisfying assumptions (3), with the spectrum  $\Sigma$  of  $H^{\omega}$  purely absolutely continuous and satisfy assumptions (1, 2). Then there exist unique  $\xi(n) \in \Psi$ , a unique partition  $\mathcal{O}_{\xi(n)}^{\pm}$  of  $\mathcal{O}_{\xi(n)}$  into disjoint subsets and unique measures  $\nu_{2,n}$  supported on  $\mathcal{O}_{\xi(n)}$  such that the  $M$ -functions have the following representation,

$$M_{\pm, n}^{\omega}(\lambda) = -\frac{1}{2} g_{\lambda}^{\omega}(n, n) \pm \int k(\lambda, x) d(\nu_{2,n}^{+} - \nu_{2,n}^{-}) \quad (63)$$

with  $\nu_{2,n}^{\pm} = \nu_{2,n}|_{\mathcal{O}_{\xi(n)}^{\pm}}$  given by

$$\nu_{2,n}^{\pm}(\xi_i(n)) = \frac{\sqrt{R(\xi_i(n))}}{\sqrt{R(\xi_i(n)) + R(\xi_{i+1}(n))}} \quad \text{for } \xi_i(n) \in \mathcal{O}_{\xi(n)}^{\pm}.$$

1967). We fix a  $\omega$  and work with it in the following, so that the superscripts are dropped. We note that for each  $n$  the real part of the Green functions  $g_\lambda(n, n)$  is zero in interior of the spectrum and it is real and increasing in the gaps  $I_i$ . Therefore for each  $n$  we get a point  $\xi(n) \in \Psi$  as in the proposition (2.2) and conclude that  $g_\lambda(n, n) = h_{\xi(n)}(\lambda)$ ,  $h$  as in proposition (2.2). By lemma (5.1) we have that  $0 < \text{Im } M_{+,n} = \text{Im } M_{-,n} < \infty$  in the interior of  $\Sigma$ . Therefore in the Herglotz representation for the  $M$ -functions, the singular part can only be supported in the set  $\partial\bar{\Sigma} \cup S_{\xi(n)}$ . This set being closed and countable, the singular part can only be pure point. We now use the relations, coming from  $(a+b)^2 - (a-b)^2 = 4ab$ , ( $F$  is defined so that we need not write long expressions later)

$$F(\lambda) \equiv (M_{+,n} - M_{-,n} - \lambda + b_n)^2 = g_\lambda(n, n)^{-2} + 4a_n^2 \frac{g_\lambda(n+1, n+1)}{g_\lambda(n, n)}$$

and the product representations for  $h_{\xi(n)}$ ,  $h_{\xi(n+1)}$  to conclude that the right hand side is an analytic function in  $\mathbb{C} \setminus (\mathcal{E} \cup S_{\xi(n)})$ . Therefore the left hand side is a meromorphic function on  $\mathbb{C}$ , with possibly essential singularities at the points of  $\mathcal{E}$ . We first rule out the set of points of  $S_{\xi(n)} \cap \Sigma$  from being singularities. If there is a singularity at any of the points in this set, then the product representations, will show that the right hand side in the above equation has a simple pole at this point, while the left hand side has a double pole giving a contradiction. This leaves us with only the set  $\mathcal{E} \cup \mathcal{O}_{\xi(n)}$ . We rule out the possibility of essential singularities at the points of  $\mathcal{E}$  by estimating the right hand side on the sequence  $R_N$  of rectangles of lemma (3.3), using the product representation for  $h_{\xi(n)}$ ,  $h_{\xi(n+1)}$  and the bounds of lemma (3.1), to show that the limits

$$\lim_{N \rightarrow \infty} \int_{R_N} F(\lambda) d\lambda = 0.$$

For a given point  $p$  in  $\mathcal{E} \setminus \mathcal{E}_1$ , we again do the estimates for  $(\lambda - p)^m F(\lambda)$ , for each positive integer  $m$  to show that there the negative terms in the Laurent series expansion for  $F$  at these points vanish. Similar method shows that none of the points of  $\mathcal{E}_1$  is a pole of any order.

The sum and total of all this analysis is that the singular parts of  $v^\pm$  can only be supported on  $\mathcal{O}_{\xi(n)}$ . The relation  $M_+(M_- + \lambda - b_n)(\lambda) = -a_n^2 g_\lambda(n+1, n+1)/g_\lambda(n, n)$  shows that a point of  $\xi_i(n) \in I_i$  can be an eigen value of only one of  $H_{\pm, n}$ , since the right hand side has a simple pole at  $\xi_i(n)$ . Therefore specifying them, which we can do in principle if we know  $H$ , will give us a partition, of  $\mathcal{O}_{\xi(n)}$  stated in the proposition. Then the measures  $v_{2,n}^\pm$  can be obtained uniquely as

$$v_{2,n}^\pm = v|_{\mathcal{O}_{\xi(n)}^\pm}$$

and the expression for  $v_{2,n}$  is clear from the product representation in  $\Pi$ , of the meromorphic function  $g_\lambda(n, n)$  coming from proposition (2.3), by a computation of the residues.  $\square$

*Lemma 5.6. Assume that  $\mathbb{P}$  satisfies assumptions (3) and the spectrum of  $H^\omega$  satisfies assumptions (1, 2). Then for each  $n$  the  $M$  functions  $M_{\pm, n}(\lambda)$  and  $(M_{\pm, n}(\lambda))^{-1}$  are uniformly bounded on the set of rectangles  $R_N$  defined in the lemma 3.3 Moreover the*

*Proof.* We consider  $\mathbb{P}$  satisfying assumptions (3) such that the corresponding random Jacobi matrix satisfies assumptions (1, 2). Consider a  $\omega$  in the support of  $\mathbb{P}$  fixed. We use the product representation of (14) for  $g_\lambda(n, n)$ , and the expressions of (63) for the  $M$ -functions. Then the bounds of lemma (3.1), prove the lemma for  $M_{\pm, n}(\lambda)$ . On the other hand for the inverses of the  $m$  functions, we use the relations of (55) on the portion of  $R_N$  in the resolvent of  $H^\pm$  and extend the bounds to all of  $R_N$  in the spectrum since the bounds are uniform in  $R_N \setminus \mathbb{R}$ . For fixed  $n$  these bounds will depend on  $1/a_n^2$ . As for the derivatives we consider the second term on the right hand side of (63). The summability of the derivatives follows from the expressions for  $v_{2, n}$  coming from the last lemma, the lower bound  $s_i$  on the distance of  $R_N$  to  $I_i$ , bounds of lemma (3.1) and assumptions (2). Regarding the boundedness of the derivative of the inverse of the Green function, we note that it is a meromorphic function in  $\mathbb{C} \setminus \Sigma$  and the derivative is given by

$$\begin{aligned} \frac{d}{d\lambda} g_\lambda(n, n)^{-1} &= \frac{1}{2} g_\lambda(n, n)^{-1} \left( \frac{1}{(\lambda - \tau_0)^2} + \frac{1}{(\lambda - \tau_\infty)^2} \right. \\ &\quad \left. + \sum_i \frac{\tau_{2i-1} - \xi_i(n)}{(\lambda - \tau_{2i-1})(\lambda - \xi_i(n))} + \frac{\tau_{2i} - \xi_i(n)}{(\lambda - \tau_{2i})(\lambda - \xi_i(n))} \right) \end{aligned}$$

The bounds of lemma (3.1) show that apriori the derivative is bounded on  $R_N \cap [\mathbb{C} \setminus \Sigma]$ , and with a uniform bound since the distance of  $R_N$  to  $I_i$  is bounded below by  $s_i$  and by assumption (2),  $q_i/s_i$  is summable. Hence even the lim sup of the derivative along any sequence on the rectangle approaching the real axis, both from  $\mathbb{C}^+$  and from  $\mathbb{C}^-$  is uniformly bounded. From the construction of  $r_N$  it is clear that the uniform bounds are valid on  $r_N$  also.  $\square$

Now we are ready to prove Theorem 1.1.

*Proof.* We consider a  $\omega$  fixed and drop the superscript in the following. We can check from the definitions of  $M_{\pm, n}$  written in terms of the solutions  $u_\pm(n)$ , and the product representations for the Green functions valid under our assumptions, that

$$\begin{aligned} \prod_{i=0}^{n-1} [-a_i^{-2}] [M_{+, i}(\lambda)] [M_{-, i}(\lambda) + \lambda - b_i] &= \frac{u_-(n)u_+(n)}{u_-(0)u_+(0)} \\ &= \frac{g_\lambda(n, n)}{g_\lambda(0, 0)} \\ &= \prod_{k=1}^{\infty} \frac{(\lambda - \xi_k(n))}{(\lambda - \xi_k(0))}. \end{aligned} \quad (64)$$

This equation shows that the product on the left hand side has zeros and poles at the points of  $S_{\xi(n)} \cup S_{\xi(0)}$  and exactly one of  $\Pi M_{+, i}$ ,  $\Pi(M_{-, i} + \lambda - b_i)$  will have a pole in the set  $S_{\xi(0)} \cap \mathcal{O}$ . On the other hand the representation for  $M_{\pm, i}$  coming from (17)

$$M_{+,i} = \frac{1}{2} \left[ \int k(\lambda, x) d(v_{2,i}^+ - v_{2,i}^-) + \lambda - b_i - g_\lambda(i, i)^{-1} \right] \\ - M_{-,i} - \lambda + b_i = \frac{1}{2} \left[ \int k(\lambda, x) d(v_{2,i}^+ - v_{2,i}^-) + \lambda - b_i + g_\lambda(i, i)^{-1} \right].$$

From the above equations, it is clear that both  $M_+$  and  $-M_- - \lambda + b_i$  are the two branches of the same meromorphic function  $\phi_i$  when lifted to the Riemann surface  $\mathcal{R}$ . Consider the meromorphic function,  $\phi(n)$ , on  $\mathcal{R}$  given by

$$\phi(n) = \prod_{i=0}^{n-1} \phi_i.$$

This function has simple zeros in  $\lambda^{-1}(S_{\xi(n)})$  and simple poles in  $\lambda^{-1}(S_{\xi(0)})$  and by the previous arguments, there is only one zero and one pole in each of  $\lambda^{-1}(\bar{I}_i)$ ,  $i = 1, 2, \dots$ . Let us denote these points as  $\xi_i^*(n)$  and  $\xi_i^*(0)$ . As for the points in  $\lambda^{-1}(\infty)$ , the asymptotic behaviour of  $M_{+,i} \approx -1/\lambda$  and  $-M_{-,i} - \lambda + b_i \approx -\lambda$  shows that  $\phi_i$  has a pole and a zero at each point of  $\lambda^{-1}(\infty)$  of order 1. Therefore the divisor of the meromorphic differential  $d\phi(n)/\phi(n)$  is given by  $D = \sum_{i=1}^{\infty} [\xi_i^*(n) - \xi_i^*(0)] + n[\infty_1 - \infty_2]$ . It follows from lemma (5.6) that the meromorphic functions  $\phi_i$  and hence their product

$$\phi(n) = \prod_{i=0}^{n-1} \phi_i$$

satisfy the assumptions of proposition (4.3) since  $\phi_i(p)$  agrees with  $M_{\pm,i}(\lambda(p))$  for  $p \in \mathcal{R}^{\pm} \cap \lambda^{-1}(\Pi)$ . Therefore the Abel-Jacobi map  $\mathcal{A}$  corresponding to the point  $\xi^*(n)$  has the image satisfying the assumptions of theorem (4.9). This shows that each of the points  $\xi_i^*(n)$ , hence their projections  $\xi_i(n)$  to  $\bar{I}_i$ , are almost periodic for each  $i$ . Now the Trace formulae, together with the uniform, in  $n$ , convergence of the sums in the Trace formulae give the almost periodicity of  $a_n^2 + a_{n-1}^2$  and  $b_n$ . To show the almost periodicity of  $a_n^2 - a_{n-1}^2$ , we compute this difference to be  $\int v_{2,n}^+ - \int v_{2,n}^-$ , by looking at the  $1/\lambda$  term in the asymptotic expansion of  $M_+ - M_-$  using the proposition (5.5). We note that this difference is nothing but the sum of the residues of the meromorphic function  $\phi_n$  (the suffix  $n$  is correct, we emphasize this) at its poles  $\xi_i^*$  in  $\lambda^{-1}(\bar{I}_i)$ , (the preimages of the open gaps, not closed gaps). This computation gives us after changing variables to the angular coordinates,

$$\int v_{2,n}^+ - \int v_{2,n}^- = \sum_{i: \xi_i(n) \in I_i} \sqrt{(\xi_i^*(n) - \tau_{2i-1})(\tau_{2i} - \xi_i^*(n))} g_i(n) \\ = \sum_{i: \xi_i(n) \in I_i} l_i \sin(\theta_i(n)) \cos(\theta_i(n)) g_i(n)$$

where we have taken,

$$\theta_j(n) = \mathcal{A}_j^{-1}(nc + d)$$

$$\xi_j^*(n) = \tau_{2j-1} + l_j \sin^2(\theta_j(n))$$

$$g_i(n) = \sqrt{(\tau_0 - \xi_i(n))(\tau_{\infty} - \xi_i(n))} \prod_{j \neq i} \frac{\sqrt{(\tau_{2j-1} - \xi_i(n))(\tau_{2j} - \xi_i(n))}}{(\xi_i(n) - \xi_j(n))}.$$

we have that modulo  $\pi$ ,  $\theta_i(n)$  is in  $[0, \pi/2)$  whenever  $\xi_i^*(n) \in \mathcal{R}^+ \cap \lambda^{-1}(I_i)$  and it is in  $[\pi/2, \pi)$  when  $\xi_i^*(n)$  is in  $\mathcal{R}^- \cap \lambda^{-1}(I_i)$ . We also note that the  $g_i(n)$  are positive. The almost periodicity of  $a_n^2 - a_{n-1}^2$ , follows from the above two equations, using the almost periodicity of  $\sin(\theta_i(n))$  and  $\cos(\theta_i(n))$  and those of  $\xi_j(n)$ , using the following lemmas, together with the uniform, in  $n$ , convergence of the products and sums. If two sequences  $c_n$  and  $d_n$  are almost periodic, then their sums, products are almost periodic. Their ratios are also almost periodic if the absolute values of the denominators have a strict positive lower bound. The positive square roots of an almost periodic sequence with positive entries is also almost periodic as can be seen from the proof below. Since by above  $a_n^2 + a_{n-1}^2$  and  $a_n^2 - a_{n-1}^2$  are almost periodic, we see that  $a_n^2$  is almost periodic. Now consider the positive square roots  $a_n$ . Then consider  $\varepsilon^4$  and an  $N$  such that  $|a_n^2 - a_{n+N}^2| < \varepsilon^4$ . For this  $N$ , consider the collection of points  $n$  such that  $a_{n+N}^2 < \varepsilon^2$ . Then for these points we have

$$|a_n - a_{n+N}| = |\sqrt{(a_n^2 - a_{n+N}^2) + a_{n+N}^2} - \sqrt{a_{n+N}^2}| \leq \sqrt{\varepsilon^4 + \varepsilon^2} + \varepsilon < 3\varepsilon.$$

for the set of points where  $a_{n+N}^2 > \varepsilon^2$ , we use the relation

$$|a_n - a_{n+N}| = \left| \frac{(a_n^2 - a_{n+N}^2)}{a_n + a_{n+N}} \right| \leq \varepsilon^2 < \varepsilon$$

using the positivity of  $a_n$ . □

Finally we remark that the equation of motion (55) for the  $m$ -functions shows that (55) does not pause in the gaps as  $n$  varies. If there is an  $n$  such that  $\zeta_i(n) = \zeta_i(n+1)$ , then it must be that these two points are eigen values of different half-line problems. We also note that the Jacobi matrices with  $a_n \equiv \text{constant}$  satisfy the condition that  $v_{2,n}^+ - v_{2,n}^- = 0$  for each  $n$ .

#### Appendix A Herglotz representation theorems.

In this appendix we collect some of the standard theorems on Herglotz functions, and state most of them without proof. We refer to (4) for the definition of the kernels  $K(\lambda, x)$  and  $k$ . The following theorems are standard and can be found in Kotani [18], Simon [35], Craig [8] or in the appendix of Figotin–Pastur [11].

**Theorem A.1.** (Herglotz representation). *Let  $F$  be Herglotz, then there are  $a \geq 0, b \in \mathbb{R}$ , such that the representation*

$$F(z) = az + b + \frac{1}{\pi} \int K(\lambda, x) d\mu(x) \tag{65}$$

*is valid, with  $\mu$  a positive borel measure satisfying  $\int 1/(1+x^2) d\mu(x) < \infty$ .*

**Theorem A.2.** *Let  $F$  and  $G$  be Herglotz functions, then the following are valid*

*The non tangential limits  $F(x+i0)$  exist finitely almost everywhere with respect to Lebesgue measure.*

*If  $F(x+i0) = G(x+i0)$ , on a set of positive Lebesgue measure then  $F \equiv G$ .*

*If  $\text{Im } F(x+i0) = 0$  a.e. on an open interval  $I$ , then  $F$  is analytic on  $I$ .*

4. If the measure  $\mu$  representing  $F$ , has compact support, then it is finite and the representation theorem for  $F$  becomes

$$F(\lambda) = a\lambda + b + \frac{1}{\pi} \int k(\lambda, x) d\mu(x).$$

5. The absolutely continuous part of the representing measure  $\mu$  can be recovered from  $F$  by the following formula. Let  $B$  be any Borel set,

$$\mu(B) = \lim_{\varepsilon \rightarrow 0} \int_B \operatorname{Im} F(x + i\varepsilon) dx$$

in particular  $\mu_{ac}$  has density  $\operatorname{Im} F(x + i0)$ .

6. The singular part of  $\mu$  is supported on  $\{x: \lim_{\varepsilon \rightarrow 0} F(x + i\varepsilon) = \infty\}$ .

*Proof.* The statement (4) follows from the trivial estimate that when the support of  $\mu$  is bounded say,  $\sup \{|x|: x \in \operatorname{supp} \mu\} = a$ , then  $1/1 + x^2 \leq 1/1 + a^2$ . Then we can absorb the quantity  $\int (x/1 + x^2) d\mu(x)$  into  $b$ , hence the stated representation. The proof for the rest can be found in the works quoted above.  $\square$

The next theorem, on the existence of zeros of a class of continuous functions, is from Deimlig [16]. Consider a bounded open set  $I$  of  $\mathbb{R}^n$  with  $\partial I$  denoting its boundary. The topological degree of a  $C^1$  map  $f$  at a point  $y$  not in the image of the boundary is defined as

$$d(f, I, y) = \sum_{y \in f^{-1}(y)} \operatorname{sgn} J_f(x), \quad y \in \mathbb{R}^n \setminus f(\partial I \cup S_f).$$

where  $S_f$  is the set of singular values of  $f$  and  $J_f$  the Jacobian. This definition is extended for  $C^2$  maps and for continuous maps it is defined in terms of  $C^2$  maps close to it, we refer the reader to the definition 2.2 of [16] for an exact definition. It follows from the definition of the degree that the identity map has degree one. The degree of a continuous map is useful in determining if it can take some values. As an application of the theory of degrees we have the following theorem.

**Theorem A.3.** Let  $I_k = [c_k, d_k]$ ;  $k = 1, 2, \dots, n$  be intervals in the real line. Define  $I = \prod_{k=1}^n I_k$ . Let  $F: I \rightarrow \mathbb{R}^n$  be a continuous function such that  $F_k(x_1, \dots, c_k, \dots, x_n) < 0$  and  $F_k(x_1, \dots, d_k, \dots, x_n) > 0 \forall k = 1, \dots, n$ , where  $F_k(\cdot)$  denotes the  $k$ -th coordinate of  $F$ . Then  $F$  has a zero in the interior of  $I$ .

*Proof.* Without loss of generality we can assume that  $c_k < 0$  and  $d_k > 0 \forall k$ . The straight line homotopy  $x \rightarrow tF(x) + (1-t)x$ ;  $0 \leq t \leq 1$  gives a homotopy between  $F$  and the identity map. The conditions  $F(x_1, \dots, c_k, \dots, x_n) < 0$  and  $F(x_1, \dots, d_k, \dots, x_n) > 0$  ensure that throughout the homotopic deformation no point in the boundary of  $I$  goes to zero. Hence the conditions of (d3) of Theorem 3.1 in [16] are valid showing that the topological degree of  $F$  for the point 0 in  $\mathbb{R}^n$  is the same as that of the identity map which is 1. This implies by (d4) of Theorem 3.1 in [16] that  $F$  has a zero in the interior of  $I$ .  $\square$

## acknowledgements

We thank Madhav Nori, S Nag, P N Srikanth, S Sastry and V S Sunder for several discussions during the course of this work and the referee of [3] for extensive comments and suggestions. We also would like to thank the Indian Academy of Sciences for funding the workshop at Kodaikanal, and the Indian Statistical Institute, Bangalore for the invitation to one of us (MK) where parts of this paper was written.

## References

- [1] Akhiezer N I, *The classical moment problem* (London: Oliver and Boyd) (1965)
- [2] Anand J Antony and Krishna M, Almost periodicity of some random Jacobi matrices *Proc. Indian Acad. Sci. (Math. Sci.)* **102** 3 (1992) 175–188
- [3] Anand J Antony and Krishna M, Inverse spectral theory for random Jacobi matrices (unpublished preprint) (1993)
- [4] Carmona R, *Springer lecture notes in Mathematics*, **1180** (1986)
- [5] Carmona R and Kotani S, Inverse spectral theory for random Jacobi matrices, *J. Stat. Phys.* **46** 5/6 (1989) 1091–1114
- [6] Carmona R and Lacroix J, *Spectral theory of random Schrödinger operators* (Stuttgart: Birkhauser) (1990)
- [7] Cycone H L, Froese R, Kirsh W and Simon B, Schrödinger operators with application to quantum mechanics and global geometry, in *Texts and Monographs in Physics* (New York: Springer-Verlag, 1987)
- [8] Craig W, Trace formulas for Schrödinger operators, *Commun. Math. Phys.* **126** (1989) 379–407
- [9] Dubrovin B A, Matveev U B and Novikov S P, Non-linear equations of Kdv type, finite zone linear operators and Abelian varieties, *Russ. Math. Surv.* **31** 1 (1976) 59–146
- [10] Farkas H M and Kra I, *Riemann surfaces* (Berlin: Springer Verlag) (1980)
- [11] Figotin A and Pastur L, Spectra of random and almost periodic operators, *Grundlehren der mathematischen Wissenschaften* 297 (New York: Springer Verlag) (1992)
- [12] Gesztesy F, Holden H, Simon B and Zhao Z, Trace formulae and inverse spectral theory for Schrödinger operators, *Bull. AMS* **29** 2 (1993) 250–255
- [13] Gesztesy F and Simon B, The xi function (preprint)
- [14] Kac M and Van Moerbeke P, On some periodic Toda lattice, *Proc. Natl. Acad. Sci. USA* **72**, (1975) 1627–1629
- [15] Kac M and Van Moerbeke P, The solution of the periodic Toda lattice, *Proc. Natl. Acad. Sci. USA* **72** (1975) 2879–2880
- [16] Deimling K, *Nonlinear Functional Analysis* (New York: Springer Verlag) (1984)
- [17] Knill O, Isospectral deformations of random Jacobi matrices, *Commun. Math. Phys.* **151** 2 (1993) 403–426
- [18] Kotani S, One dimensional random Schrödinger operators and Herglotz functions, *Proc. Taniguchi Symp.* (ed.) S A Katata (1987) 443–452
- [19] Kotani S, On the inverse problems for random Schrödinger operators, *AMS Series of Contemporary Mathematics* **41** (1985) 267–280
- [20] Kotani S, Link between periodic potentials and random potentials in one dimensional random Schrödinger operators, in *Proc. Int. Conf. Diff. Equ. Math. Phys.* (1986)
- [21] I W Knowles and Y Saito (eds) (Berlin: Springer Verlag) 1987
- [22] Kotani S, Absolutely continuous spectrum of one dimensional random Schrödinger operators and Hamiltonian systems in *Proc. Jap. USSR probability Symp.* (1986)
- [23] Kotani S, Jacobi matrices with random potentials taking finitely many values, *Rev. Math. Phys.* **1** 1 (1987) 123–124
- [24] Kotani S and Krishna M, Almost periodicity of some random potentials, *J. Funct. Anal.* **78** (1989) 390–405
- [25] Kra I, *Automorphic forms and Kleinian groups* (Reading, Massachusetts: W A Benjamin) (1972)

- [25] Kunz H and Souillard B, Sur les spectre des opérateurs aux différences finies élatoires, *Commun. Math. Phys.* **78** (1980) 201–246
- [26] Lang S, *Analysis II*, (London: Addison-Wesley) 1969
- [27] Levitan B M, On the closure of finite zone potentials, *Math. USSR-Sbornik* **51** (1985) 67–89
- [28] Levitan B M, An inverse problem for the Sturm-Liouville operator in the case of finite-zone and infinite-zone potentials, *Trudy Moskov Mat. Obsch.* **45** (1982) 3–36, English Translation *Moscow Math. Soc.* **1** (1984)
- [29] Levitan B M, Almost periodicity of infinite zone potentials, *Math USSR Izetsija* **182** (1982) 249–274
- [30] Moser J, *Integrable Hamiltonian systems and spectral theory* (Pisa: Lezioni Fermiani) (1981)
- [31] McKean H P and Van Moerbecke P, The spectrum of Hills operator, *Invent. Math.* **30** (1975) 217–274
- [32] McKean H P and Trubowitz E, Hills operator and hyper-elliptic function theory in the presence of infinitely many branch points, *Commun. Pure Appl. Math.* **29** (1976) 143–226
- [33] Pastur L A, Spectral properties of Disordered systems in the one body approximation, *Commun. Math. Phys.* **75** (1980) 179–196
- [34] Rajaram Bhat B V and Parthasarathy K R, Generalized harmonic oscillators in quantum probability, *Seminaire de probabilites-XXV, Springer Lecture Notes in Mathematics* **1485** (1991) 39–51
- [35] Simon B, Kotani theory for one dimensional random Jacobi matrices, *Commun. Math. Phys.* **89** (1983) 227
- [36] Simon B, Spectral analysis of rank one perturbations and applications, Lectures delivered at Vancouver summer school, Caltech preprint (1993)
- [37] Toda M, Theory of non-linear lattices, *Solid State Series* (New York: Springer-Verlag) **20** (1989)
- [38] Trubowitz E, The inverse problem for periodic potentials, *Comm. Pure Appl. Math.* **30** (1977) 321–337
- [39] Van Moerbecke P, The spectrum of Jacobi matrices, *Invent. Math.* **37** (1976) 45–81

### Note added in proof

In the summer of 1994, we came across the work *Infinite dimensional Jacobi inversion problem, almost periodic Jacobi matrices with homogeneous spectrum and Hardy classes of character automorphic functions*, Preprint 1994 of M Sodin and P Yuditskii. In this work Sodin–Yuditskii prove the almost periodicity of reflection less Jacobi matrices with compact homogeneous spectra, which includes cantor like spectra.



## Spectral shift function and trace formula

KALYAN B SINHA and A N MOHAPATRA

Indian Statistical Institute 7, S.J.S. Sansanwal Marg, New Delhi 110016, India

**Abstract.** The complete proofs of Krein's theorem on the spectral shift function and the trace formula are given for a pair of self-adjoint operators such that either (i) their difference is trace-class or (ii) the difference of their resolvents is trace-class. The proofs, essentially due to Krein, is based on Herglotz's theorem on the boundary value of the analytic functions whose imaginary part is non-negative on the upper half plane, and an almost optimal class of functions are obtained for which the trace formula is valid. Also an alternative method based on Weyl-von Neumann's theorem for self-adjoint operators, avoiding the complex function theory and inspired by Voiculescu's work, is given for the first case. Furthermore, some applications of the spectral shift function have been discussed.

**Keywords.** Spectral shift function; trace formula; Krein's theorem.

## Introduction

Krein's spectral shift function and associated trace formulas [9, 18, 19, 20, 30] have been of considerable interest as an abstract mathematical statement as well as for various applications. The original proof of Krein (see for example [20]) uses analytic function theory and we use the same in §2 and §4. Voiculescu [28] gave a proof of the trace formula without using function theory for the case of bounded self-adjoint operators. We extend this method in §3 to a pair of arbitrary self-adjoint operators whose difference is trace-class. In the appendix we collect some of the necessary results from analytic function theory without proof as well as the definition and some properties of the perturbation determinant. Section 5 deals with some applications. In this article,  $\mathcal{H}$  will denote the Hilbert space we work in;  $\mathcal{B}(\mathcal{H})$ ,  $\mathcal{B}_1(\mathcal{H})$  and  $\mathcal{B}_2(\mathcal{H})$  standing for the set of bounded, trace-class and Hilbert-Schmidt operators respectively. We shall often have  $H$  and  $H_0$  as a pair of self-adjoint operators in  $\mathcal{H}$  with  $\sigma(H)$ ,  $\sigma(H_0)$  their spectra;  $\rho(H)$ ,  $\rho(H_0)$  their resolvent sets with  $R_z$  and  $R_z^0$  their resolvents and  $E_\lambda$ ,  $E_\lambda^0$  the associated spectral families. The symbols  $\|\cdot\|$ ,  $\|\cdot\|_1$  and  $\|\cdot\|_2$  denote operator norm, trace norm and Hilbert-Schmidt norm respectively, while  $\text{Tr}$  will stand for the trace of a trace-class operator  $B$ . In a finite dimensional Hilbert space the problem is easy to state and prove.

**Theorem 1.1.** Let  $H$  and  $H_0$  be two self-adjoint operators in a finite dimensional Hilbert space  $\mathcal{H}$ . Then there exists a unique real-valued bounded function  $\xi$  such that

$$(i) \quad \xi(\lambda) = \text{Tr}(E_\lambda^0 - E_\lambda) \quad \lambda \in \mathbb{R}$$

(iii) for  $\varphi \in C^1(\mathbb{R})$ ,

$$\text{Tr}[\varphi(H) - \varphi(H_0)] = \int \varphi'(\lambda) \xi(\lambda) d\lambda. \quad (1.1)$$

Furthermore,  $\xi$  is a constant in every real open interval in  $\rho(H) \cap \rho(H_0)$  and has support in  $[a, b]$ , where  $a = \min\{\inf \sigma(H), \inf \sigma(H_0)\}$ ,  $b = \max\{\sup \sigma(H), \sup \sigma(H_0)\}$ .

(iv) If  $H - H_0 = \tau|g\rangle\langle g|$  with  $\tau > 0$ ,  $\|g\| = 1$  (we have used Dirac notation for rank one operators), then  $\xi$  is a  $\{0, 1\}$ -valued function. More precisely,  $\xi(\lambda) = \sum_{j=1}^r \chi_{\Delta_j}(\lambda)$  for  $r$  disjoint intervals  $\Delta_j \subseteq \mathbb{R}$ ,  $1 \leq r \leq n$ .

*Proof.* In fact, we define  $\xi$  by (i). Then  $\xi$  is a bounded real-valued function with the stated support property. We only prove (iii). Given the support of  $\xi$ , the integral on the right hand side of (1.1) is over a finite interval only. By functional calculus,  $\text{Tr}[\varphi(H) - \varphi(H_0)] = -\int \varphi(\lambda) d\xi(\lambda) = -\xi(\lambda) \varphi(\lambda)|_{-\infty}^{\infty} + \int \varphi'(\lambda) \xi(\lambda) d\lambda$ , and the result follows. That  $\xi$  is constant in every open real interval in  $\rho(H) \cap \rho(H_0)$  is a consequence of the definition of the spectral families.

Let  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$  and  $\mu_1 \leq \mu_2 \leq \dots \leq \mu_n$  be the eigenvalues of  $H_0$  and  $H$  respectively. Then since  $H - H_0$  is positive rank one, it is a consequence of the minimax principle for eigenvalues [14] that  $\lambda_j \leq \mu_j \leq \lambda_{j+1}$  ( $1 \leq j \leq n-1$ ) and  $\lambda_n \leq \mu_n$ . From this it is clear that if  $\lambda_j = \lambda_{j+1} = \lambda_{j+2} = \dots = \lambda_{j+s}$  for some  $j$  and  $s$ , then  $\mu_j = \mu_{j+1} = \dots = \mu_{j+s-1} = \lambda_j$ . This and the expression (i) for  $\xi$  leads to:

$$\xi(\lambda) = \sum_{j=1}^n \chi_{[\lambda_j, \mu_j]}(\lambda),$$

where we set  $\chi_{[\lambda_j, \mu_j]}(\lambda) = 0$  for those  $j$  for which  $\lambda_j = \mu_j$ . ■

The next corollary follows easily from the theorem.

## COROLLARY 1.2

For  $t \in \mathbb{R}$ ,  $\text{Tr}(e^{itH} - e^{itH_0}) = it \int e^{it\lambda} \xi(\lambda) d\lambda$ .

In an infinite dimensional Hilbert space, the relation  $\xi(\lambda) = \text{Tr}(E_\lambda^0 - E_\lambda)$  will not make sense in general because  $E_\lambda^0 - E_\lambda$  may not be trace-class. Next we give a counter example due to Krein [18] where  $H - H_0$  is rank one and yet  $E_\lambda - E_\lambda^0$  is not trace-class.

*Counter example.* Let  $\mathcal{H} = L^2[0, \infty)$  and  $L = -\frac{d^2}{dx^2}$  be the differential operator with

$D(L) = C_0^\infty(0, \infty)$ . It is known that  $L$  has several self-adjoint extensions depending upon the boundary conditions on the corresponding differential equation. Of them choose two, namely  $h_0$  and  $h$  as

$$D(h_0) = \left\{ f \in \mathcal{H} : f \text{ and } f' \text{ are absolutely continuous, } \right. \\ \left. f'' \in \mathcal{H} \text{ and } f(0) = 0 \right\}$$

and

$$D(h) = \left\{ f \in \mathcal{H} : f \text{ and } f' \text{ are absolutely continuous, } \right. \\ \left. f'' \in \mathcal{H} \text{ and } f'(0) = 0. \right\}$$

Note that both  $h_0$  and  $h$  are positive operators. One can compute the spectral families  $F_\lambda^0$  and  $F_\lambda$  of  $h_0$  and  $h$  respectively by solving the associated ordinary differential equations and get for  $\lambda \geq 0$

$$F_\lambda^0(x, y) = \frac{2}{\pi} \int_0^{(\lambda)^{1/2}} \sin tx \sin ty dt \quad (1.2)$$

and

$$F_\lambda(x, y) = \frac{2}{\pi} \int_0^{(\lambda)^{1/2}} \cos tx \cos ty dt.$$

Let  $H_0 = (h_0 + I)^{-1}$  and  $H = (h + I)^{-1}$ . The Green's functions associated with  $(h_0 + I)^{-1}$  and  $(h + I)^{-1}$  are

$$G_0(x, y) = \begin{cases} e^{-y} \sinh x, & x \leq y \\ e^{-x} \sinh y, & x \geq y \end{cases} \quad (1.3)$$

and

$$G(x, y) = \begin{cases} e^{-y} \cosh x, & x \leq y \\ e^{-x} \cosh y, & x \geq y \end{cases}$$

respectively. Then  $H - H_0 = \frac{1}{2} |\psi\rangle \langle \psi|$ , where  $\psi(x) = \sqrt{2}e^{-x}$  so that  $\|\psi\| = 1$ . Let

$\mu = \frac{1}{1 + \lambda}$ . Then  $E_\mu^0 = I - F_\lambda^0$  and  $E_\mu = I - F_\lambda$  are the spectral families of  $H_0$  and  $H$  respectively and

$$E_\mu(x, y) - E_\mu^0(x, y) = F_\lambda^0(x, y) - F_\lambda(x, y) = (-2/\pi) \frac{\sin \sqrt{\lambda}(x+y)}{(x+y)}. \quad (1.4)$$

Note that  $E_\mu - E_\mu^0$  is not trace-class since the Hilbert-Schmidt norm

$$\begin{aligned} \iint |E_\mu(x, y) - E_\mu^0(x, y)|^2 dx dy &= (2/\pi)^2 \int_0^\infty \int_0^\infty \frac{\sin^2 \sqrt{\lambda}(x+y)}{(x+y)^2} dx dy \\ &= \begin{cases} \infty & \text{if } \lambda \neq 0, \text{ equivalently if } 0 < \mu < 1, \\ 0 & \text{if } \lambda = 0, \text{ equivalently if } \mu = 1. \end{cases} \end{aligned}$$

If  $E_\mu - E_\mu^0$  were trace class, then since its integral kernel (1.4) is continuous, we could evaluate the trace (see p. 523 of [16]) as:

$$\begin{aligned} \text{Tr}(E_\mu^0 - E_\mu) &= \int_0^\infty (E_\mu^0 - E_\mu)(x, x) dx \\ &= \frac{2}{\pi} \int_0^\infty \frac{\sin 2\sqrt{\lambda}x}{2x} dx \\ &= \frac{1}{2} \quad \text{if } 0 < \mu < 1 \text{ and} \\ &= 0 \quad \text{if } \mu = 1. \end{aligned} \quad (1.5)$$

In § 2 we shall define Krein's spectral shift function  $\xi$  when the difference of  $H$  and  $H_0$  is trace class, and in remark 2.7 (iv) it will be shown that  $\xi$  in the above example is precisely the expression (1.5) though  $E_\mu^0 - E_\mu$  is not trace class.

Here we mention some other authors who have also dealt with this subject, in particular Clancy [10] and Kuroda [17]. There is also the interesting approach of Birman and Solomyak [8] using the theory of double spectral integrals, also developed by them. They obtain a trace formula of the type (1.1) for a function class somewhat larger than what we have described in § 2-4.

## 2. Spectral shift function and trace formula: the case of trace-class perturbation

In this section we shall establish the existence of the spectral shift function and prove the trace formula (1.1) for a large class of functions under the assumption  $V \in \mathcal{B}_1$ . The following theorem concerns the first assertion, (see also [20], [25], [10]).

**Theorem 2.1** *Let  $H$  and  $H_0$  be two self-adjoint operators in  $\mathcal{H}$  such that  $V = H - H_0 \in \mathcal{B}_1$ . Let  $\Delta(z) = \det(I + VR_z^0)$ , for  $\text{Im } z \neq 0$ , be the perturbation determinant (see appendix for definition and its properties). Then there exists a unique real valued  $L^1(\mathbb{R})$ -function  $\xi$  satisfying*

- (i) 
$$\xi(\lambda) = \frac{1}{\pi} \lim_{\varepsilon \rightarrow 0+} \text{Im} \ln \Delta(\lambda + i\varepsilon) \quad (2.1)$$
- (ii) 
$$\int_{-\infty}^{\infty} |\xi(\lambda)| d\lambda \leq \|V\|_1, \quad \int_{-\infty}^{\infty} \xi(\lambda) d\lambda = \text{Tr } V$$
- (iii) 
$$\ln \Delta(z) = \int_{-\infty}^{\infty} \frac{\xi(\lambda)}{\lambda - z} d\lambda \quad \text{for } \text{Im } z \neq 0,$$
- (iv) 
$$\text{Tr}(R_z - R_z^0) = - \int_{-\infty}^{\infty} \frac{\xi(\lambda)}{(\lambda - z)^2} d\lambda \quad \text{for } \text{Im } z \neq 0,$$

*Proof.* Let  $V$  be self-adjoint, rank one, i.e.  $V = \tau|g\rangle\langle g|$ ,  $\tau \neq 0$  and  $\|g\| = 1$ . Then for  $\text{Im } z \neq 0$ ,

$$\Delta(z) = 1 + \tau(g, R_z^0 g) = 1 + \tau \int \frac{d\|E_\lambda^0 g\|^2}{\lambda - z}. \quad (2.2)$$

In the following by  $\ln$  we mean the principal branch of the logarithm function.

If  $\text{Im } z > 0$ , then by (2.2),  $\tau^{-1} \text{Im} \Delta(z) = (\text{Im } z) \int \frac{d\|E_\lambda^0 g\|^2}{|\lambda - z|^2} > 0$  or equivalently  $0 \leq (\text{sgn } \tau) \text{Im} \ln \Delta(z) < \pi$ , where  $\text{sgn } \tau = \pm 1$  according as  $\tau > 0$  or  $< 0$ . Since  $\Delta(z)$  is analytic and has no zeros there,  $G(z) \equiv (\text{sgn } \tau) \ln \Delta(z)$  is analytic in the upper half-plane  $\{z \in \mathbb{C} : \text{Im } z > 0\}$  and, by (2.2),  $|G(z)| = O\left(\frac{1}{\text{Im } z}\right)$  as  $\text{Im } z \rightarrow \infty$ . So by theorem A.1

there exists a non-negative function  $\zeta \in L^1$ , given by  $\zeta(\lambda) = \frac{1}{\pi} \lim_{\varepsilon \rightarrow 0+} \text{Im} G(\lambda + i\varepsilon)$  for almost all  $\lambda$  such that

$$(\text{sgn } \tau) \ln \Delta(z) = G(z) = \int_{-\infty}^{\infty} \frac{\zeta(\lambda)}{\lambda - z} d\lambda.$$

Set  $\xi(\lambda) = (\operatorname{sgn} \tau) \zeta(\lambda)$ . Then by theorem A.5 (i) and the above relation one obtains

$$\operatorname{Tr}(R_z - R_z^0) = -\frac{d}{dz} \ln \Delta(z) = -\int_{-\infty}^{\infty} \frac{\xi(\lambda)}{(\lambda - z)^2} d\lambda, \quad (2.3)$$

since the integral in the right hand side converges uniformly in  $z$  for  $\operatorname{Im} z \geq \delta > 0$ .

Using the resolvent identity  $R_z - R_z^0 = -R_z V R_z^0$  and the fact  $s - \lim_{y \rightarrow \infty} iy R_{iy} = s - \lim_{y \rightarrow \infty} iy R_{iy}^0 = -I$ , one gets  $y^2(R_{iy} - R_{iy}^0) \rightarrow V$  in  $\mathcal{B}_1$ -norm as  $y \rightarrow \infty$  (see lemma 8.23 of [1]). Since  $\xi \in L^1$ , an application of dominated convergence theorem then yields

$$\begin{aligned} \tau &= \operatorname{Tr} V = \lim_{y \rightarrow \infty} y^2 \operatorname{Tr}(R_{iy} - R_{iy}^0) \\ &= -\lim_{y \rightarrow \infty} \int_{-\infty}^{\infty} \frac{\xi(\lambda) y^2}{(\lambda - iy)^2} d\lambda = \int_{-\infty}^{\infty} \xi(\lambda) d\lambda. \end{aligned} \quad (2.4)$$

So by (2.4)

$$\int_{-\infty}^{\infty} |\xi(\lambda)| d\lambda = \int_{-\infty}^{\infty} \xi(\lambda) d\lambda = (\operatorname{sgn} \tau) \int_{-\infty}^{\infty} \xi(\lambda) d\lambda = (\operatorname{sgn} \tau) \tau = |\tau| = \|V\|_1.$$

Next, let  $V \in \mathcal{B}_1$  and write  $V = \sum_{j=1}^{\infty} \tau_j |g_j\rangle \langle g_j|$  with  $\sum_{j=1}^{\infty} |\tau_j| < \infty$ ,  $\|g_j\| = 1$  for each  $j$ . Set for  $k = 1, 2, \dots$ ,  $H_k = H_0 + V_k \equiv H_0 + \sum_{j=1}^k \tau_j |g_j\rangle \langle g_j|$ ,  $R_z^{(k)} = (H_k - z)^{-1}$  and  $\Delta_k(z) = \det(I + (H_k - H_{k-1}) R_z^{(k-1)})$ . Then, as before, for each  $k \geq 1$ , there exists a real valued function  $\xi_k \in L^1$  such that  $\int \xi_k(\lambda) d\lambda = \tau_k$  and  $\int |\xi_k(\lambda)| d\lambda = |\tau_k|$ . Define  $\xi(\lambda) = \sum_{k=1}^{\infty} \xi_k(\lambda)$ ; the right hand side converges in  $L^1$ -norm since  $\sum_{k=1}^{\infty} |\tau_k| < \infty$ . This also shows that  $\xi \in L^1(\mathbb{R})$ . Note that  $\xi$  is unique since each  $\xi_k$  is. Moreover

$$\int_{-\infty}^{\infty} |\xi(\lambda)| d\lambda \leq \sum_{k=1}^{\infty} \int_{-\infty}^{\infty} |\xi_k(\lambda)| d\lambda = \sum_{k=1}^{\infty} |\tau_k| = \|V\|_1,$$

and

$$\int_{-\infty}^{\infty} \xi(\lambda) d\lambda = \sum_{k=1}^{\infty} \int_{-\infty}^{\infty} \xi_k(\lambda) d\lambda = \sum_{k=1}^{\infty} \tau_k = \operatorname{Tr} V$$

This proves (ii).

By theorem A.5 (i)

$$\begin{aligned} \ln \det(I + V_k R_z^0) &= \sum_{j=1}^k \ln \Delta_j(z) \\ &= \int d\lambda \left\{ \sum_{j=1}^k \xi_j(\lambda) \right\} (\lambda - z)^{-1}. \end{aligned}$$

As  $k \rightarrow \infty$ , the right hand side converges to  $\int \frac{\xi(\lambda)}{\lambda - z} d\lambda$  for  $\operatorname{Im} z \neq 0$ , whereas by the continuity property of determinant (see theorem A.3 (ii)) the left hand side converges to  $\ln \Delta(z)$  since  $V_k \rightarrow V$  in  $\mathcal{B}_1$ -norm and  $\det(I + V_k R_z^0) \neq 0$  for  $\operatorname{Im} z \neq 0$  and all  $k$ . Thus we get

$$\ln \Delta(z) = \int_{-\infty}^{\infty} \frac{\xi(\lambda)}{\lambda - z} d\lambda. \quad (2.5)$$

which proves (iii). The property (i) follows from theorem 13 in [27].

By theorem A.5 (i) and (2.5) we have  $\text{Tr}(R_z - R_z^0) = - \int_{-\infty}^{\infty} \frac{\xi(\lambda)}{(\lambda - z)^2} d\lambda$  by differentiating inside the integral of (2.5), which is allowed since second integral converges uniformly in  $z$  for  $\text{Im } z \geq \delta > 0$ . ■

Next we study the class of functions  $\varphi$  for which  $\varphi(H) - \varphi(H_0) \in \mathcal{B}_1$  and the trace formula (1.1) holds. For example, let  $\varphi(\lambda) = e^{it\lambda}$  for fixed  $t \in \mathbb{R}$ . Note that for  $t \neq 0$

$$\frac{e^{itH} - e^{itH_0}}{it} = \frac{1}{t} e^{itH_0} \int_0^t ds e^{-isH_0} V e^{isH} \quad (2.6)$$

on  $D(H_0)$ . Since  $s \rightarrow e^{-isH_0}$ ,  $e^{isH}$  are strongly continuous and since  $V \in \mathcal{B}_1$ , it follows by lemma 8.23 of [1] that  $s \rightarrow e^{-isH_0} V e^{isH}$  is  $\mathcal{B}_1$ -continuous. This means that the Riemann integral exists in  $\mathcal{B}_1$ -norm, the left hand side of (2.6) converges to  $V$  as  $t \rightarrow 0$  in  $\mathcal{B}_1$ -norm and we have the estimate:

$$\left\| \frac{e^{itH} - e^{itH_0}}{it} \right\|_1 \leq \|V\|_1. \quad (2.7)$$

Therefore  $\varphi(H) - \varphi(H_0) \in \mathcal{B}_1$ . It is also clear that  $t \rightarrow \frac{e^{itH} - e^{itH_0}}{it}$  is a  $\mathcal{B}_1$ -continuous map.

We will prove an abstract theorem from which the trace formula for  $\varphi(\lambda) = e^{it\lambda}$  will follow. Using that the trace formula for a large class of functions will be established. We start with the following lemma.

**Lemma 2.2.** *Let  $A$  be a self-adjoint operator in  $\mathcal{H}$ , so that  $\Psi_z \equiv (A - i)(A - z)^{-1} \in \mathcal{B}(\mathcal{H})$  for  $\text{Im } z \neq 0$ . Then for any  $\varepsilon \neq 0$  and  $g \in \mathcal{H}$*

$$\begin{aligned} \text{(i)} \quad & \int_{-\infty}^{\infty} d\lambda \|(A - \lambda - i\varepsilon)^{-1} g\|^2 = \pi |\varepsilon|^{-1} \|g\|^2 \\ \text{(ii)} \quad & \int_{-\infty}^{\infty} \frac{d\lambda}{1 + \lambda^2} \|\Psi_{\lambda + i\varepsilon} g\|^2 \leq 2\pi(1 + |\varepsilon|^{-1}) \|g\|^2 \end{aligned}$$

*Proof.* Let  $\{F_\lambda\}$  be the spectral family of  $A$ . Then by functional calculus and Fubini's theorem,

$$\begin{aligned} \int_{-\infty}^{\infty} d\lambda \|(A - \lambda - i\varepsilon)^{-1} g\|^2 &= \int_{-\infty}^{\infty} d\lambda \int_{-\infty}^{\infty} \frac{d\|F_\mu g\|^2}{(\mu - \lambda)^2 + \varepsilon^2} \\ &= \int_{-\infty}^{\infty} d\|F_\mu g\|^2 \int_{-\infty}^{\infty} \frac{d\lambda}{(\mu - \lambda)^2 + \varepsilon^2} \\ &= \pi |\varepsilon|^{-1} \|g\|^2. \end{aligned}$$

Similarly,

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{d\lambda}{1 + \lambda^2} \|\Psi_{\lambda + i\varepsilon} g\|^2 &= \int_{-\infty}^{\infty} \frac{d\lambda}{1 + \lambda^2} \int_{-\infty}^{\infty} d\|F_\mu g\|^2 \frac{(\mu^2 + 1)}{(\mu - \lambda)^2 + \varepsilon^2} \\ &= \int_{-\infty}^{\infty} d\|F_\mu g\|^2 \int_{-\infty}^{\infty} \frac{(\mu^2 + 1) d\lambda}{\{(\mu - \lambda)^2 + \varepsilon^2\}(1 + \lambda^2)}, \end{aligned}$$

from which part (ii) follows since the integrand is dominated by

$$\frac{2(\mu - \lambda)^2 + 2(\lambda^2 + 1)}{\{(\mu - \lambda)^2 + \varepsilon^2\}(1 + \lambda^2)} \leq \frac{2}{1 + \lambda^2} + \frac{2}{(\mu - \lambda)^2 + \varepsilon^2}. \quad (2.8)$$

Let  $\varphi \in C(\mathbb{R})$  be bounded. For  $\varepsilon > 0$ , define

$$\varphi_\varepsilon(\lambda) = \frac{\varepsilon}{\pi} \int_{-\infty}^{\infty} \frac{\varphi(\mu)}{(\mu - \lambda)^2 + \varepsilon^2} d\mu. \quad (2.9)$$

**Remark 2.3.** If  $\varphi$  and  $\varphi_\varepsilon$  are as above, then it is easy to see that for any self-adjoint operator  $A$ ,  $\varphi(A)$  and  $\varphi_\varepsilon(A)$ , defined by functional calculus, are bounded operators and  $\varphi_\varepsilon(A)$  converges strongly to  $\varphi(A)$  as  $\varepsilon \rightarrow 0$ .

**Theorem 2.4.** Let  $H$  and  $H_0$  be two self-adjoint operators in  $\mathcal{H}$  such that  $V = H - H_0 \in \mathcal{B}_1$ . Let  $\varphi \in C(\mathbb{R})$  be bounded and  $\varphi_\varepsilon$  be given by (2.9). Then

- (i)  $\varphi_\varepsilon(H) - \varphi_\varepsilon(H_0) \in \mathcal{B}_1$ .
- (ii) If furthermore  $\varphi \in C^1(\mathbb{R})$  and  $\varphi'$  is bounded, then

$$\lim_{\varepsilon \rightarrow 0} \text{Tr} \{ \varphi_\varepsilon(H) - \varphi_\varepsilon(H_0) \} = \int \varphi'(\lambda) \xi(\lambda) d\lambda,$$

where  $\xi$  is the function given by theorem 2.1.

(iii) If also  $\varphi_\varepsilon(H) - \varphi_\varepsilon(H_0)$  converges in  $\mathcal{B}_1$ -norm as  $\varepsilon \rightarrow 0$ , then the trace formula (1.1) holds.

**Proof.** Let  $\varepsilon > 0$  be fixed. Then by (2.9)

$$\varphi_\varepsilon(H) - \varphi_\varepsilon(H_0) = \frac{1}{\pi} \int d\lambda \varphi(\lambda) \text{Im}(R_{\lambda + i\varepsilon} - R_{\lambda + i\varepsilon}^0). \quad (2.10)$$

Since  $\varphi$  is bounded, for part (i) it suffices to show that  $T_\pm(\lambda) \equiv \|R_{\lambda \pm i\varepsilon} - R_{\lambda \pm i\varepsilon}^0\|_1 \in L^1(\mathbb{R}; d\lambda)$ . We give the proof of positive sign only, the other case being similar.

Let  $V = \sum_{k=1}^{\infty} \tau_k |g_k\rangle \langle g_k|$  with  $\sum_{k=1}^{\infty} |\tau_k| < \infty$  and  $\{g_k\}$  an orthonormal set. Then the resolvent identity  $R_z - R_z^0 = -R_z V R_z^0$  leads to the estimate:

$$\begin{aligned} \|R_z - R_z^0\|_1 &\leq \sum_{k=1}^{\infty} |\tau_k| \|R_z g_k\| \|R_z^0 g_k\| \\ &= \sum_{k=1}^{\infty} |\tau_k| \|R_z g_k\| \|R_z^0 g_k\|. \end{aligned}$$

Setting  $z = \lambda + i\varepsilon$  and integrating both the sides of the above inequality with respect to  $\lambda$  and using Schwarz's inequality and lemma 2.2 (i) we obtain

$$\int_{-\infty}^{\infty} d\lambda \|R_{\lambda + i\varepsilon} - R_{\lambda + i\varepsilon}^0\|_1$$

$$\leq \sum_{k=1}^{\infty} |\tau_k| \left\{ \int_{-\infty}^{\infty} d\lambda \|R_{\lambda - i\varepsilon}^0 g_k\|^2 \right\} \left\{ \int_{-\infty}^{\infty} d\lambda \|R_{\lambda + i\varepsilon} g_k\|^2 \right\} \\ = (\pi/\varepsilon) \|V\|_1.$$

Assume  $\varphi'$  to be bounded and continuous. Then by (2.10), theorem 2.1 (iii) and integration by parts we get

$$\begin{aligned} \text{Tr}\{\varphi_\varepsilon(H) - \varphi_\varepsilon(H_0)\} &= \frac{1}{\pi} \int_{-\infty}^{\infty} d\lambda \varphi(\lambda) \text{Tr}\{\text{Im}(R_{\lambda + i\varepsilon} - R_{\lambda + i\varepsilon}^0)\} \\ &= -\frac{1}{\pi} \int_{-\infty}^{\infty} d\lambda \varphi(\lambda) \left\{ \text{Im} \int_{-\infty}^{\infty} \frac{\xi(\mu) d\mu}{(\mu - \lambda - i\varepsilon)^2} \right\} \\ &= -\frac{1}{\pi} \int_{-\infty}^{\infty} d\lambda \varphi(\lambda) \frac{d}{d\lambda} \left\{ \varepsilon \int_{-\infty}^{\infty} \frac{\xi(\mu)}{(\mu - \lambda)^2 + \varepsilon^2} d\mu \right\} \\ &= -\frac{\varepsilon}{\pi} \left[ \varphi(\lambda) \int_{-\infty}^{\infty} \frac{\xi(\mu)}{(\mu - \lambda)^2 + \varepsilon^2} d\mu \right]_{\lambda=-\infty}^{\lambda=+\infty} \\ &\quad + \frac{\varepsilon}{\pi} \int_{-\infty}^{\infty} d\lambda \varphi'(\lambda) \int_{-\infty}^{\infty} \frac{\xi(\mu)}{(\mu - \lambda)^2 + \varepsilon^2} d\mu. \end{aligned}$$

Note that  $\varphi(\lambda) \int_{-\infty}^{\infty} \frac{\xi(\mu)}{(\mu - \lambda)^2 + \varepsilon^2} d\mu \rightarrow 0$  as  $|\lambda| \rightarrow \infty$  by dominated convergence theorem. Hence the boundary terms in the above vanishes, and thus

$$\begin{aligned} \text{Tr}\{\varphi_\varepsilon(H) - \varphi_\varepsilon(H_0)\} &= \frac{\varepsilon}{\pi} \int_{-\infty}^{\infty} d\lambda \varphi'(\lambda) \int_{-\infty}^{\infty} \frac{\xi(\mu)}{(\mu - \lambda)^2 + \varepsilon^2} d\mu \\ &= \int_{-\infty}^{\infty} d\mu \xi(\mu) \left\{ \frac{\varepsilon}{\pi} \int_{-\infty}^{\infty} \frac{\varphi'(\lambda)}{(\mu - \lambda)^2 + \varepsilon^2} d\lambda \right\} \\ &\equiv \int_{-\infty}^{\infty} d\mu \xi(\mu) \Phi_\varepsilon(\mu), \end{aligned}$$

which converges to  $\int_{-\infty}^{\infty} \varphi'(\mu) \xi(\mu) d\mu$  by dominated convergence theorem since  $\Phi_\varepsilon(\mu) \rightarrow \varphi'(\mu)$  for every  $\mu$  (see theorem 13 of [27]) as  $\varepsilon \rightarrow 0$  and is bounded by  $\left( \sup_{\lambda \in \mathbb{R}} |\varphi'(\lambda)| \right) \frac{\varepsilon}{\pi} \int_{-\infty}^{\infty} \frac{d\lambda}{(\mu - \lambda)^2 + \varepsilon^2} = \sup_{\lambda \in \mathbb{R}} |\varphi'(\lambda)|$ . This proves (ii).

If  $\varphi_\varepsilon(H) - \varphi_\varepsilon(H_0)$  converges in  $\mathcal{B}_1$  as  $\varepsilon \rightarrow 0$ , then it converges to  $\varphi(H) - \varphi(H_0)$  since  $\varphi_\varepsilon(H)$  and  $\varphi_\varepsilon(H_0)$  converge strongly to  $\varphi(H)$  and  $\varphi(H_0)$  respectively. Hence the trace formula follows from part (ii). ■

As a corollary of this theorem we obtain

#### COROLLARY 2.5

Let  $H$ ,  $H_0$  and  $\xi$  be as in theorem 2.1. Then for each  $t \in \mathbb{R}$ ,  $e^{itH} - e^{itH_0} \in \mathcal{B}_1$  and

$$\text{Tr}(e^{itH} - e^{itH_0}) = it \int_{-\infty}^{\infty} e^{it\lambda} \xi(\lambda) d\lambda. \quad (2.11)$$



of. Let  $t \in \mathbb{R}$  be fixed. Then the first part follows from (2.7). Set  $\varphi(\lambda) = e^{it\lambda}$ . Then (2.9), given in (2.9), can be easily computed:  $\varphi_\varepsilon(\lambda) = e^{-\varepsilon|\lambda|} e^{it\lambda}$ . Clearly  $\varphi$  satisfies the hypotheses of theorem 2.4, and  $\varphi_\varepsilon(H) - \varphi_\varepsilon(H_0) = e^{-\varepsilon|t|}(e^{itH} - e^{itH_0})$  converges in  $\mathcal{B}_1$  to  $e^{itH} - e^{itH_0}$  as  $\varepsilon \rightarrow 0$ . Thus the result follows from theorem 2.4 (iii). ■

A function  $\varphi$  on  $\mathbb{R}$  is said to be in Krein class  $\mathcal{K}$  if  $\varphi$  is given by

$$\varphi(\lambda) = \int_{-\infty}^{\infty} \frac{e^{it\lambda} - 1}{it} \nu(dt) + C \quad (2.12)$$

for some constant  $C$  and complex measure  $\nu$  on  $\mathbb{R}$ . Note that such a function is necessarily continuously differentiable and the derivative is the Fourier transform of the measure  $\nu$  i.e.,

$$\varphi'(\lambda) = \int e^{it\lambda} \nu(dt).$$

It is also worth observing that  $\varphi(H)$  and  $\varphi(H_0)$  are not necessarily bounded operators (see remark 2.7 (ii)) though defined on  $D(H) = D(H_0)$ .

**Theorem 2.6.** Let  $H, H_0$  and  $\xi$  be as in theorem 2.1, and  $\varphi \in \mathcal{K}$ . Then  $\varphi(H) - \varphi(H_0) \in \mathcal{B}_1$

$$\text{Tr}\{\varphi(H) - \varphi(H_0)\} = \int_{-\infty}^{\infty} \varphi'(\lambda) \xi(\lambda) d\lambda.$$

Proof. By functional calculus

$$\varphi(H) - \varphi(H_0) = \int_{-\infty}^{\infty} \frac{e^{itH} - e^{itH_0}}{it} \nu(dt), \quad (2.13)$$

as we have observed that by the discussion following the proof of theorem 2.1. By estimate (2.7), the integral in (2.13) exists as a  $\mathcal{B}_1$ -valued Bochner integral (page 1 of [5]). It also follows that

$$\|\varphi(H) - \varphi(H_0)\|_1 \leq \|V\|_1 \int_{-\infty}^{\infty} |\nu|(dt) < \infty,$$

so we have by (2.11)

$$\begin{aligned} \text{Tr}\{\varphi(H) - \varphi(H_0)\} &= \int_{-\infty}^{\infty} \frac{\nu(dt)}{it} \text{Tr}(e^{itH} - e^{itH_0}) \\ &= \int_{-\infty}^{\infty} \frac{\nu(dt)}{it} \int_{-\infty}^{\infty} e^{it\lambda} \xi(\lambda) d\lambda \\ &= \int_{-\infty}^{\infty} d\lambda \xi(\lambda) \int_{-\infty}^{\infty} e^{it\lambda} \nu(dt) \\ &= \int_{-\infty}^{\infty} \varphi'(\lambda) \xi(\lambda) d\lambda \end{aligned}$$

In the above the change in the order of integration is justified since

$$\int_{-\infty}^{\infty} d\lambda |\xi(\lambda)| \int_{-\infty}^{\infty} |v|(dt) < \infty.$$

*Remark 2.7.* (i) If  $\text{supp } v$  does not contain 0, then  $\varphi \in \mathcal{K}$  is a bounded function. On the other hand if we set  $v(dt) = \delta(t)dt$  or  $= \frac{it}{2\pi} \hat{\xi}(t)dt$  (with  $C=0$  or  $= \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{\xi}(t)dt$  respectively), where  $\xi \in \mathcal{S}(\mathbb{R})$ , the Schwartz class of smooth functions of rapid decrease, and  $\hat{\xi}$  its Fourier transform, then  $\varphi(\lambda) = \lambda$  or  $= \xi(\lambda)$  respectively.

(ii) Since  $\left\| \frac{1}{it}(e^{itH} - I)f \right\| \leq \|Hf\|$ ,  $\varphi(H)$  is well defined on  $D(H) = D(H_0)$  for  $\varphi \in \mathcal{K}$ , and by (2.7),  $\varphi(H) - \varphi(H_0)$  can be extended to whole of  $\mathcal{K}$  as a trace-class operator.

(iii) Let  $J$  be a real open interval in  $\rho(H) \cap \rho(H_0)$ , and let  $\varphi \in C_0^\infty(J)$ , the class of smooth functions with compact support in  $J$ . Then by functional calculus  $\varphi(H) - \varphi(H_0) = 0$ , and hence by the trace formula in theorem 2.6,

$$\int \varphi'(\lambda) \xi(\lambda) d\lambda = 0 \quad (2.14)$$

for all  $\varphi \in C_0^\infty(J) \subseteq \mathcal{K}$ . Since  $\xi \in L^1$ , it follows that  $\xi \in L_{loc}^1(\mathbb{R})$ , and hence  $\xi$  can be thought of as a distribution on  $J$ . Then the equation (2.14) can be viewed as  $\langle \xi', \varphi \rangle = 0$  for all  $\varphi \in C_0^\infty(J)$  where  $\xi'$  is the distributional derivative. By a standard theorem in the theory of distributions (see p. 105 of [12])  $\xi$  is constant in  $J$ . Furthermore if  $J$  contains a neighbourhood of either  $+\infty$  or  $-\infty$ , then  $\xi = 0$  a.e. on  $J$  since by theorem 2.1, (ii)  $\int_{-\infty}^{\infty} |\xi(\lambda)| d\lambda \leq \|V\|_1 < \infty$ . Thus,  $\text{supp } \xi$  lies in the interval  $[a, b]$ ,

where  $a = \min \{ \inf \sigma(H_0), \inf \sigma(H) \}$  and  $b = \max \{ \sup \sigma(H_0), \sup \sigma(H) \}$ .

(iv) We would like to go back once again to the counterexample in § 1 and observe the curious fact that the "formal expression" for  $\text{Tr}(E_\mu^0 - E_\mu)$  coincides exactly with the boundary value of the argument of the perturbation determinant in this case.

Using the notation of section 1, we see that  $(E_\mu^0 \psi, \psi) = 2 \int_0^{(\lambda)^{1/2}} \alpha^2 (1 + \alpha^2)^{-1} d\alpha$  with  $\mu = (1 + \lambda)^{-1}$ . Thus for  $\text{Im } z \neq 0$ ,

$$\begin{aligned} \Delta(z) &= 1 + \frac{1}{2} \int_0^1 \frac{d(E_\mu^0 \psi, \psi)}{\mu - z} \\ &= 1 + \frac{1}{\pi} \int_0^1 \left( \frac{1 - \mu}{\mu} \right)^{1/2} (\mu - z)^{-1} d\mu. \end{aligned}$$

The last integral can be evaluated using the calculus of residues (see for example [26]) and this yields

$$\begin{aligned}\xi(\mu) &= \lim_{\varepsilon \rightarrow 0+} \frac{1}{\pi} \arg \Delta(\mu + i\varepsilon) \\ &= \frac{1}{2} \quad \text{if } 0 < \mu < 1 \\ &= 0 \quad \text{if } \mu \notin [0, 1].\end{aligned}$$

(v) In some applications [13],  $H - H_0$  may not be trace-class but  $e^{-tH} - e^{-tH_0}$  is trace-class for some  $t > 0$  (with  $H$  and  $H_0 \geq 0$ ). Such a case can be treated by the results of this section. Let  $A = e^{-H}$ ,  $B = e^{-H_0}$  so that  $0 \leq A$ ,  $B \leq I$  and assume that  $B \in \mathcal{B}_1$ . Then by theorem 2.1 and remark 2.7 (iii) one has  $\eta \in L^1[0, 1]$  such that  $\int_0^1 \eta(\mu) d\mu \leq \|e^{-H} - e^{-H_0}\|_1$  and

$$\int_0^1 \eta(\mu) d\mu = \text{Tr}(e^{-H} - e^{-H_0}). \quad (2.16)$$

Setting  $\mu = e^{-\lambda}$  ( $0 \leq \lambda < \infty$ ) and  $\xi(\lambda) = -\eta(e^{-\lambda})$  in (2.16) we have

$$\begin{aligned}\int_0^\infty |\xi(\lambda)| e^{-\lambda} d\lambda &\leq \|e^{-H} - e^{-H_0}\|_1 \\ -\int_0^\infty \xi(\lambda) e^{-\lambda} d\lambda &= \text{Tr}(e^{-H} - e^{-H_0}).\end{aligned}$$

Consider the function  $g(\mu) = \mu^t$  for  $\mu \in [0, 1]$  and  $t > 2$ . Since  $g$  is  $C^2[0, 1]$  function and  $g''' \in L^1[0, 1]$ , we can find a function  $G \in \mathcal{H}$  such that  $G(\mu) = g(\mu)$  for all  $\mu \in [0, 1]$ .  
S

$$\begin{aligned}\text{Tr}(e^{-tH} - e^{-tH_0}) &= \text{Tr}[(e^{-H})^t - (e^{-H_0})^t] \\ &= \int_0^1 \frac{d}{d\mu}(\mu^t) \eta(\mu) d\mu = -t \int_0^\infty e^{-t\lambda} \xi(\lambda) d\lambda.\end{aligned} \quad (2.17)$$

(vi) More generally one can use the formula (5.7) and the invariance principle of scattering theory to derive the trace formula. As in [23] a real valued function  $\psi$  on an open subset of  $\mathbb{R}$ , is said to be *admissible* if  $J = \cup_1^N J_n$  where  $J_n = (\alpha_n, \beta_n)$  are disjoint,  $N$  finite or infinite, and (i)  $\psi'' \in L_{loc}^1(J)$ , (ii)  $\psi' > 0$  or  $< 0$  on each interval  $J_n$ . Then one has

**Theorem 2.8.** [23] (invariance principle). *Let  $\psi$  be an admissible function on  $J$ ,  $H$  and  $H_0$  be selfadjoint operators such that  $\sigma(H), \sigma(H_0) \subset \bar{J}$  and that at each boundary point of  $J$  either  $\psi$  has a finite limit or both  $H$  and  $H_0$  do not have point spectrum at that point. Suppose furthermore  $H - H_0 \in \mathcal{B}_1$ . Then  $\Omega_\pm(\psi(H), \psi(H_0))$  exist, are complete and*

where  $J_1$  (respectively  $J_2$ ) is the union of those intervals on which  $\psi' > 0$  (respectively  $\psi' < 0$ ).

Then using (5.4)–(5.7) one gets in the spectral representation of  $H_0$ :

$$\xi(\lambda; H, H_0) = \text{sgn}(\psi'(\lambda)) \cdot \xi(\psi(\lambda); \psi(H), \psi(H_0)). \quad (2.18)$$

The relation (2.18) can be turned around to give a definition of  $\xi(\lambda; H, H_0)$  when  $H - H_0$  is not trace-class but rather  $\psi(H) - \psi(H_0)$  is trace-class. It is also clear then that  $\psi(\lambda) = e^{-t\lambda} (\lambda \geq 0)$  is an admissible function for every  $t > 0$ , and this gives the example in (v).

Now, if  $\psi$  is an admissible function such that  $\psi(H) - \psi(H_0) \in \mathcal{B}_1$  and  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$  be such that  $\varphi \circ \psi^{-1}$  (note that  $\psi^{-1}$  exists) is again an admissible function, then writing  $\varphi(H) - \varphi(H_0) = \varphi \circ \psi^{-1}(\psi(H)) - \varphi \circ \psi^{-1}(\psi(H_0))$ , one has formally (by using (2.18) and a change of variable  $\mu = \psi(\lambda)$ )

$$\begin{aligned} \text{Tr}[\varphi(H) - \varphi(H_0)] &= \int \xi(\mu; \psi(H), \psi(H_0)) (\varphi \circ \psi^{-1})'(\mu) d\mu \\ &= \int \xi(\lambda; H, H_0) (\varphi \circ \psi^{-1})'(\psi(\lambda)) \text{sgn} \psi'(\lambda) |\psi'(\lambda)| d\lambda \\ &= \int \xi(\lambda; H, H_0) \varphi'(\lambda) d\lambda. \end{aligned}$$

### 3. An alternative proof of the trace formula

Here we give a functional analytic proof (following Voiculescu [28]) of the results in §1. The strategy is to reduce the computation of  $\text{Tr}(e^{itH} - e^{itH_0})$  to that of  $\text{Tr}(e^{itH_m} - e^{itH_{0,m}})$  for suitable finite dimensional approximations  $H_m$  and  $H_{0,m}$  of  $H$  and  $H_0$  respectively and then apply theorem 1.1. We begin with a few lemmas which are extensions of Weyl-von Neumann result (see lemma 2.2 of p. 523 of [16]).

*Lemma 3.1.* *Let  $A$  be a self-adjoint operator in  $\mathcal{H}$ ,  $f \in \mathcal{H}$  and  $\varepsilon > 0$ ,  $K$  a compact set in  $\mathbb{R}$ . Then there exist a projection  $P$  of finite rank in  $\mathcal{H}$  such that*

- (i)  $\|(I - P)e^{itA}P\|_2 < \varepsilon$  uniformly for  $t \in K$
- (ii)  $\|(I - P)f\| < \varepsilon$ .

*Proof.* Let  $F(\cdot)$  be the spectral measure associated with the self-adjoint operator  $A$ . Choose  $a > 0$  such that  $\|(I - F(-a, a])f\| < \varepsilon$ . For each positive integer  $n$  and  $1 \leq k \leq n$ , set  $F_k = F\left(\frac{2k-2-n}{n}a, \frac{2k-n}{n}a\right]$  and note that  $F_k F_j = \delta_{kj} F_j$ ,  $\sum_{k=1}^n F_k = F(-a, a]$ . We also set  $g_k = \begin{cases} F_k f / \|F_k f\| & \text{if } F_k f \neq 0 \\ 0 & \text{otherwise.} \end{cases}$  Then  $g_k \in D(A)$  and  $Ag_k \in F_k \mathcal{H}$ . Let  $P$  be the projection on to the sub-space generated by  $\{g_1, \dots, g_n\}$  so that  $\dim P\mathcal{H} \leq n$ . With  $\lambda_k = \frac{2k-n-1}{n}a$ , it is easy to verify that

$$\|(A - \lambda_k)g_k\|^2 = \int_{[(2k-n-2)/n]a}^{[(2k-n)/n]a} (\lambda - \lambda_k)^2 d\|F(\lambda)g_k\|^2 \leq (a/n)^2,$$

$$\|(I - P)APu\|^2 = \left\| \sum_{k=1}^n (u, g_k)(I - P)Ag_k \right\|^2 \leq (a/n)^2 \|u\|^2$$

$u \in \mathcal{H}$  and hence

$$\|(I - P)AP\|_2 \leq a/\sqrt{n}.$$

$$\alpha(t) \equiv \|(I - P)e^{itA}P\|_2 = \|(I - P)(e^{itA} - I)P\|_2$$

$$= \|(I - P) \int_0^t e^{isA} iA ds P\|_2$$

$$\leq \int_0^t \{ \|(I - P)e^{isA}P\|_2 \|AP\| + \|(I - P)e^{isA}(I - P)\| \|(I - P)AP\|_2 \} ds$$

$$\leq 2a \int_0^t \alpha(s) ds + Ta/\sqrt{n} \quad (3.1)$$

$|t| < T$ . We can solve this Gronwall-type inequality (3.1) to conclude that

$$\alpha(t) \leq (Tae^{2aT})/\sqrt{n} \leq (Tae^{2aT})/\sqrt{n}.$$

On the other hand,  $(I - P)F(-a, a]f = \sum_{k=1}^n \|F_k f\| (I - P)g_k = 0$  so that  $\|(I - P)f\| = \|(I - P)(I - F(-a, a])f\| < \varepsilon$ . ■

**Lemma 3.2** Let  $H$  and  $H_0$  be two selfadjoint operators such that  $V \equiv H - H_0$  is positive and of rank one. Set  $V = \tau |g\rangle\langle g|$  with  $\tau > 0$  and  $\|g\| = 1$ . Then given any  $\varepsilon > 0$ , there exists a projection  $P$  of finite rank such that for all  $t$  with  $|t| < T$ .

- (i)  $\|(I - P)g\| < \varepsilon$ ,  $\|(I - P)e^{itH_0}P\|_2 < \varepsilon$ ,  $\|(I - P)e^{itH_0}g\| < 2\varepsilon$ ,
- (ii)  $\|(I - P)HP\|_2 < \varepsilon(1 + \tau)$ ,  $\|(e^{itH} - e^{itH_0})(I - P)\|_1 < 2T\tau\varepsilon$ ,
- (iii)  $\|P(e^{itH_0} - e^{itPH_0P})P\|_1 < \varepsilon^2 T$ ,  $\|P(e^{itH} - e^{itPH_0P})P\|_1 \leq \varepsilon^2 T(1 + \tau)$ ,
- (iv)  $|\text{Tr}(e^{itH} - e^{itH_0}) - \text{Tr}\{P(e^{itPH_0P} - e^{itPH_0P})P\}| \leq T\varepsilon[\tau(4 + \varepsilon) + 2\varepsilon]$ .

**Proof.** Given  $\varepsilon > 0$ ,  $g$  and  $H_0$ , we construct  $P$  as in lemma 3.1 so that the first two conclusions of (i) are satisfied. The third one follows from the first two trivially. Since  $(I - P)HP = (I - P)H_0P + \tau|(I - P)g\rangle\langle Pg|$ , the first part of (ii) follows from the estimate of lemma 3.1 and (i). Now

$$\begin{aligned} \|(e^{itH_0} - e^{itH})(I - P)\|_1 &= \|\tau \int_0^t |e^{i(t-s)H}g\rangle\langle (I - P)e^{isH_0}g| ds\|_1 \\ &\leq \tau \int_0^{|t|} \|(I - P)e^{isH_0}g\| ds \leq 2T\tau\varepsilon. \end{aligned}$$

This easily leads to the fact that

$$\|(1-P)(e^{itH} - e^{itH_0})P\|_1 < 2T\tau\varepsilon.$$

For (iii) we observe that

$$\begin{aligned} \|P(e^{itH_0} - e^{itPH_0P})P\|_1 &\leq \int_0^{|t|} \|Pe^{i(t-s)H_0}(I-P)H_0e^{isPH_0P}P\|_1 ds \\ &\leq \int_0^{|t|} \|Pe^{i(t-s)H_0}(I-P)\|_2 \|(I-P)H_0P\|_2 ds \end{aligned}$$

and an application of (i) and the estimate in lemma 3.1 gives the result. A similar computation and the estimates in (ii) give the second result in (iii).

Finally since  $\|e^{itH} - e^{itH_0}\|_1 \leq \tau \int_0^{|t|} \|e^{i(t-s)H}g\|_1 \|e^{isH_0}g\|_1 ds \leq \tau T$ , it follows that  $e^{itH} - e^{itH_0} \in \mathcal{B}_1$  and we have by (ii) and (iii)

$$\begin{aligned} &|\operatorname{Tr}(e^{itH} - e^{itH_0}) - \operatorname{Tr}\{P(e^{itPH_0P} - e^{itPH_0P})P\}| \\ &\leq \|P(e^{itH} - e^{itPH_0P})P\|_1 + \|P(e^{itH_0} - e^{itPH_0P})P\|_1 \\ &\quad + \|(e^{itH} - e^{itH_0})(I-P)\|_1 + \|(I-P)(e^{itH} - e^{itH_0})P\|_1 \\ &\leq T\varepsilon[\tau(4+\varepsilon) + 2\varepsilon]. \end{aligned}$$

Now we are ready to prove the main theorem of this section.

**Theorem 3.3.** Let  $H_0$  be a selfadjoint operator and  $H = H_0 + V$  with  $V$  self-adjoint trace-class. Then there exists a unique real-valued function  $\xi$  in  $L^1(\mathbb{R})$  such that

- (i)  $\operatorname{Tr}(e^{itH} - e^{itH_0}) = it \int e^{i\lambda} \xi(\lambda) d\lambda$ ,
- (ii)  $\int \xi(\lambda) d\lambda = \operatorname{Tr} V$ ,  $\int |\xi(\lambda)| d\lambda \leq \|V\|_1$ ,
- (iii) for every function  $\varphi \in \mathcal{K}$  (defined in (2.12)),  $\varphi(H) - \varphi(H_0) \in \mathcal{B}_1$  and

$$\operatorname{Tr}(\varphi(H) - \varphi(H_0)) = \int_{-\infty}^{\infty} \varphi'(\lambda) \xi(\lambda) d\lambda,$$

- (iv) the function  $\lambda \rightarrow (\lambda - z)^{-1}$  (with  $\operatorname{Im} z \neq 0$ ) belongs to the class  $\mathcal{K}$  and hence

$$\operatorname{Tr}(R_z - R_z^0) = - \int (\lambda - z)^{-2} \xi(\lambda) d\lambda.$$

*Proof.* At first we let  $V \equiv \tau|g\rangle\langle g|$ ,  $\tau > 0$ ,  $\|g\| = 1$ . Then we rephrase the conclusion (iv) of lemma 3.2 as: there exists a sequence  $P_m$  of finite rank projections such that  $P_m g \rightarrow g$  strongly and

$$\operatorname{Tr}(e^{itH} - e^{itH_0}) = \lim_{m \rightarrow \infty} \operatorname{Tr}[P_m(e^{itH_m} - e^{itH_{0,m}})P_m], \quad (3.2)$$

where  $H_{0,m} = P_m H_0 P_m$  and  $H_m = P_m H P_m$ , and the convergence is uniform in  $t$  for

$< T$ . Note that by construction  $P_m \mathcal{H} \subseteq D(H_0) = D(H)$  and hence both  $H_{0,m}$  and  $P_m$  are self-adjoint operators in the finite-dimensional space  $P_m \mathcal{H}$ . Next we use Theorem 1.1 (iv) and corollary 1.2 in the right hand side of (3.2) to get a  $\{0, 1\}$ -valued  $\lambda$ -function  $\xi_m$  such that

$$\mathrm{Tr}[P_m(e^{itH_m} - e^{itH_{0,m}})P_m] = it \int_{-\infty}^{\infty} e^{it\lambda} \xi_m(\lambda) d\lambda. \quad (3.3)$$

ence

$$\mathrm{Tr}(e^{itH} - e^{itH_0}) = it \lim_{m \rightarrow \infty} \int_{-\infty}^{\infty} e^{it\lambda} \xi_m(\lambda) d\lambda, \quad (3.4)$$

the convergence being uniform in  $t$ . It is easy to see from (3.3) that

$$\begin{aligned} \int \xi_m(\lambda) d\lambda &= \lim_{t \rightarrow 0} \frac{1}{it} \mathrm{Tr}[P_m(e^{itH_m} - e^{itH_{0,m}})P_m] \\ &= \lim_{t \rightarrow 0} \frac{1}{t} \int_0^t \mathrm{Tr}[P_m e^{i(t-s)H_m} P_m V P_m e^{isH_{0,m}} P_m] ds \\ &= \mathrm{Tr} P_m V P_m, \end{aligned} \quad (3.5)$$

since  $t \rightarrow e^{itH_m}$ ,  $e^{itH_{0,m}}$  are norm continuous in  $P_m \mathcal{H}$  and  $P_m V P_m$  is rank one. Thus  $\int \xi_m(\lambda) d\lambda = \tau \|P_m g\|^2 > \tau(1-\varepsilon)^2$  by lemma 3.2 (i) and setting  $\mu_m(\Delta) = (\tau \|P_m g\|^2)^{-1} \int_{\Delta} \xi_m(\lambda) d\lambda$  for every Borel set  $\Delta \subseteq \mathbb{R}$ , we have a family  $\{\mu_m\}$  of probability measures (3.5). Also note that by (3.4) the family  $\{\hat{\mu}_m(t)\}$  of their Fourier transforms converges

$\hat{\mu}(t) = \frac{1}{it\tau} \mathrm{Tr}(e^{itH} - e^{itH_0})$  uniformly in  $t$  in compact sets in  $\mathbb{R} \setminus \{0\}$ . On the other

and  $\hat{\mu}_m(0) = (\tau \|P_m g\|^2)^{-1} \int \xi_m(\lambda) d\lambda = 1$  for all  $m$  and a calculation identical to that (3.5) shows that  $\lim_{t \rightarrow 0} \hat{\mu}(t) = \tau^{-1} \mathrm{Tr} V = 1 \equiv \hat{\mu}(0)$ , by definition. Thus by Levy-Cramer continuity theorem [22], there exists a probability measure  $\mu$  on  $\mathbb{R}$  such that  $\mu_m \rightarrow \mu$  weakly i.e.  $\int \varphi(\lambda) d\mu_m(\lambda) \rightarrow \int \varphi(\lambda) d\mu(\lambda)$  as  $m \rightarrow \infty$  for every bounded continuous function  $\varphi$ .

Let  $\Delta = (a, b] \subseteq \mathbb{R}$  and let  $\{\varphi_n\}$  be a sequence of smooth functions of support in  $[-\frac{1}{n}, b + \frac{1}{n}]$  such that  $0 \leq \varphi_n \leq 1$  and  $\|\chi_{\Delta} - \varphi_n\|_1 \rightarrow 0$  as  $n \rightarrow \infty$  where  $\chi_{\Delta}$  is the indicator function of  $\Delta$ . Choosing a subsequence if necessary and using the bounded convergence theorem, we have

$$\lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \int \varphi_n(\lambda) d\mu_m(\lambda) = \lim_{n \rightarrow \infty} \int \varphi_n(\lambda) d\mu(\lambda) = \mu(\Delta).$$

us

$$\begin{aligned} \mu(\Delta) &= \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \frac{1}{\tau \|P_m g\|^2} \int \varphi_n(\lambda) \xi_m(\lambda) d\lambda \\ &= \frac{1}{\tau} \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \int \varphi_n(\lambda) \xi_m(\lambda) d\lambda \\ &\leq \tau^{-1} \lim_{n \rightarrow \infty} (b - a + 2/n) = \tau^{-1} (b - a), \end{aligned}$$

since  $0 \leq \xi_m(\lambda) \leq 1$  for all  $m$  and all  $\lambda$ . This shows that  $\mu$  is absolutely continuous and we set  $\xi(\lambda) = \tau \frac{d\mu(\lambda)}{d\lambda}$ . Then  $\xi$  is a non-negative  $L^1$  function and we have that  $\hat{\mu}(t) = \int e^{it\lambda} d\mu(\lambda) = \tau^{-1} \int e^{it\lambda} \xi(\lambda) d\lambda$  and hence

$$\text{Tr}(e^{itH} - e^{itH_0}) = it \int e^{it\lambda} \xi(\lambda) d\lambda. \quad (3.6)$$

Also dividing both sides of (3.6) by  $it$  and taking limit  $t \rightarrow 0$  as in (3.5) we conclude that

$$\int \xi(\lambda) d\lambda = \text{Tr } V = \tau \geq 0.$$

If  $V$  is rank one and negative, then we interchange the role of  $H$  and  $H_0$  and write  $H_0 = H - V$  with  $-V$  rank one and positive and obtain as above a non-negative  $L^1$ -function  $\eta$  such that

$$\text{Tr}(e^{itH_0} - e^{itH}) = it \int e^{it\lambda} \eta(\lambda) d\lambda$$

and

$$\int \eta(\lambda) d\lambda = \text{Tr}(-V) = -\text{Tr } V \geq 0.$$

Defining  $\xi(\lambda) = -\eta(\lambda)$ , we get that relation (3.6) is valid for all  $V$  rank one with some real-valued  $L^1$ -function  $\xi$ .

Now let  $V \in \mathcal{B}_1$  and let  $V = \sum_{k=1}^{\infty} \tau_k |g_k\rangle \langle g_k|$  be its canonical decomposition with  $\|V\|_1 = \sum_{k=1}^{\infty} |\tau_k| < \infty$  and  $\|g_k\| = 1$ . We write for  $k=1, 2, \dots$ ,  $H_k = H_0 + \sum_{j=1}^k \tau_j |g_j\rangle \langle g_j|$ ,  $|g_j\rangle = H_{k-1} + \tau_k |g_k\rangle \langle g_k|$ . Then we have a real valued  $L^1$ -function  $\xi_k$  such that

$$\begin{aligned} \text{Tr}(e^{itH_k} - e^{itH_{k-1}}) &= it \int e^{it\lambda} \xi_k(\lambda) d\lambda, \\ \int \xi_k(\lambda) d\lambda &= \tau_k \quad \text{and} \quad \int |\xi_k(\lambda)| d\lambda = |\tau_k|. \end{aligned}$$

Set  $\xi(\lambda) = \sum_{k=1}^{\infty} \xi_k(\lambda)$ , then by the above relations  $\int |\xi(\lambda)| d\lambda \leq \sum_{k=1}^{\infty} \int |\xi_k(\lambda)| d\lambda = \sum_{k=1}^{\infty} |\tau_k| = \|V\|_1$  and thus  $\xi$  is a real-valued  $L^1$  function and  $\sum_{k=1}^{\infty} \xi_k$  converges in  $L^1$ -norm. Therefore

$$\begin{aligned} it \int e^{it\lambda} \xi(\lambda) d\lambda &= it \sum_{k=1}^{\infty} \int e^{it\lambda} \xi_k(\lambda) d\lambda \\ &= \sum_{k=1}^{\infty} \text{Tr}(e^{itH_k} - e^{itH_{k-1}}) \\ &= \lim_{k \rightarrow \infty} \text{Tr}(e^{itH_k} - e^{itH_0}) \\ &= \text{Tr}(e^{itH} - e^{itH_0}), \end{aligned}$$



since

$$\begin{aligned}\|e^{itH} - e^{itH_k}\|_1 &= \|i \sum_{j=k+1}^{\infty} \tau_j \int_0^t |e^{i(t-s)H} g_j\rangle \langle e^{-isH_k} g_j| ds\|_1 \\ &\leq |t| \sum_{j=k+1}^{\infty} |\tau_j| \rightarrow 0 \quad \text{as } k \rightarrow \infty.\end{aligned}$$

Also

$$\int \xi(\lambda) d\lambda = \sum_{k=1}^{\infty} \int \xi_k(\lambda) d\lambda = \sum_{k=1}^{\infty} \tau_k = \text{Tr } V.$$

This completes the proof of (i) and (ii). The proof of part (iii) follows as in theorem 2.6.

For (iv) we just note that  $(\lambda - z)^{-1} + z^{-1} = \int \frac{e^{it\lambda} - 1}{it} \nu(dt)$  with  $\nu(dt) = -t\chi_{\mp}(t)e^{-izt}dt$ , according as  $\text{Im } z \geq 0$ , where  $\chi_{\pm}$  are the indicator functions of the intervals  $[0, \infty)$  and  $(-\infty, 0]$  respectively. ■

#### 4. The trace formula: the case when the difference of resolvents is trace-class

In this section we shall follow essentially the methods of §2, but for the case when the perturbation  $V$  is not necessarily of trace class but is such that the difference of resolvents  $R_z - R_z^0$  is trace-class for some  $z \in \rho(H) \cap \rho(H_0)$ . It is not difficult to see that if  $R_z - R_z^0 \in \mathcal{B}_1$  for some such  $z$ , then it is so for all such  $z$  and hence we shall, in this section, take  $z = i$  as the reference point and assume that  $R_i - R_i^0 \in \mathcal{B}_1$ . Also it is worth noting that in  $L^2(\mathbb{R}^3)$ , if  $H_0 = -\Delta$  and  $V$  is the multiplication operator by a function  $V \in L^2(\mathbb{R}^3) \cap L^1(\mathbb{R}^3)$ , then  $R_i - R_i^0 \in \mathcal{B}_1$  (see p. 546 of [16]).

We set  $U_0 = \frac{H_0 + i}{H_0 - i} = I + 2iR_i^0$  and  $U = \frac{H + i}{H - i} = I + 2iR_i$  so that  $U - U_0 = 2i(R_i - R_i^0) \in \mathcal{B}_1$ . If we also set  $U - U_0 = U_0 T$  then it is clear that  $T$  is a normal trace-class operator and  $I + T$  is unitary. Let  $T = \sum_{j=1}^{\infty} \tau_j |g_j\rangle \langle g_j|$  be the canonical decomposition for  $T$  with  $\|g_j\| = 1$  and  $1 + \tau_j = \exp(i\theta_j)$ ,  $-\pi < \theta_j \leq \pi$ . Then it follows that

$$\begin{aligned}\sum_{j=1}^{\infty} |\theta_j| &= \sum_{j=1}^{\infty} \left| e^{-i\theta_j/2} \left[ \frac{\theta_j/2}{\sin(\theta_j/2)} \right] \tau_j \right| \leq \frac{\pi}{2} \sum_{j=1}^{\infty} |\tau_j| \\ &= \frac{\pi}{2} \|T\|_1 < \infty\end{aligned}\tag{4.1}$$

since  $\left| \frac{\sin \theta}{\theta} \right| \geq 2/\pi$  for  $0 \leq \theta \leq \pi/2$ . Thus the determinant

$$\Delta(\omega) \equiv \det[(U - \omega)(U_0 - \omega)^{-1}] = \det[I + U_0 T(U_0 - \omega)^{-1}]$$

is analytic for  $|\omega| < 1$  and has no zeroes there (see theorem A.5 (ii)). Next we obtain the Krein's spectral shift function in this case essentially following the same route as

a real-valued function  $\xi$  on  $\mathbb{R}$  such that

- (i)  $\xi(\lambda)(1 + \lambda^2)^{-1} \in L^1(\mathbb{R})$ ,
- (ii)  $\int_{-\infty}^{\infty} |\xi(\lambda)|(1 + \lambda^2)^{-1} d\lambda \leq (\pi/4) \|T\|_1$  and  $\int_{-\infty}^{\infty} \xi(\lambda)(1 + \lambda^2)^{-1} d\lambda = \frac{-i}{2} \ln \det(I + T)$ ,
- (iii)  $\xi(\lambda) \equiv \eta(\alpha) = \frac{1}{\pi} \lim_{\rho \uparrow 1} \operatorname{Im} \ln \left[ \exp \left( -\frac{i}{2} \sum_{j=1}^{\infty} \theta_j \right) \Delta(\rho e^{i\alpha}) \right]$ , with  $e^{i\alpha} = (\lambda + i)(\lambda - i)^{-1}$ ,
- (iv)  $\operatorname{Tr}(R_z - R_z^0) = - \int_{-\infty}^{\infty} (\lambda - z)^{-2} \xi(\lambda) d\lambda$  for  $\operatorname{Im} z \neq 0$ .

Furthermore,  $\xi$  is unique up to an additive constant function.

*Proof.* At first let  $T$  be rank one i.e.  $T = \tau |g\rangle \langle g|$  with  $1 + \tau = e^{i\theta}$  ( $-\pi < \theta \leq \pi$ ),  $\|g\| = 1$ . Then

$$\begin{aligned} \Delta(\omega) &= \det[I + \tau |U_0 g\rangle \langle (U_0^* - \bar{\omega})^{-1} g|] = 1 + \tau(g, U_0(U_0 - \omega)^{-1} g) \\ &= e^{i\theta/2} \left[ \cos(\theta/2) + i \sin(\theta/2) \int_{-\pi}^{\pi} \frac{e^{i\alpha} + \omega}{e^{i\alpha} - \omega} d\|F_0(\alpha)g\|^2 \right], \end{aligned} \quad (4.2)$$

where  $F_0$  is the spectral family of the unitary operator  $U_0$ . Thus

$$\operatorname{Im}[e^{-i\theta/2} \Delta(\omega)] = \sin(\theta/2) \int_{-\pi}^{\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(\alpha - \beta)} d\|F_0(\alpha)g\|^2, \quad (4.3)$$

where we have set  $\omega = \rho \exp(i\beta)$ ,  $0 \leq \rho < 1$ ; and hence  $(\operatorname{sgn} \theta) \cdot \operatorname{Im}[e^{-i\theta/2} \Delta(\omega)] \geq 0$  or equivalently  $0 \leq (\operatorname{sgn} \theta) \cdot \operatorname{Im} \ln[e^{-i\theta/2} \Delta(\omega)] \leq \pi$ . Since  $\Delta(\omega)$  is analytic in the interior or the exterior of the unit circle and has no zeroes there,  $\ln \Delta(\omega)$  is also analytic there. Therefore by theorem A.2 there exists a real-valued  $L^1[-\pi, \pi]$  function  $\eta(\alpha)$  such that

$$\ln[e^{-i\theta/2} \Delta(\omega)] = \operatorname{Re} \ln[e^{-i\theta/2} \Delta(0)] + \frac{i}{2} \int_{-\pi}^{\pi} \frac{e^{i\alpha} + \omega}{e^{i\alpha} - \omega} \eta(\alpha) d\alpha.$$

Now  $\Delta(0) = \det(UU_0^{-1}) = \det[U_0(I + T)U_0^{-1}] = \det(I + T) = e^{i\theta}$  by theorem A.3 (v), so that  $\operatorname{Re} \ln[e^{-i\theta/2} \Delta(0)] = 0$  and we have

$$\ln \Delta(\omega) = i\theta/2 + \frac{i}{2} \int_{-\pi}^{\pi} \frac{e^{i\alpha} + \omega}{e^{i\alpha} - \omega} \eta(\alpha) d\alpha. \quad (4.4)$$

We also know from the theorem A.2 that  $\eta(\alpha) = \frac{1}{\pi} \lim_{\rho \uparrow 1} \operatorname{Im} \ln[e^{-i\theta/2} \Delta(\rho e^{i\alpha})]$  and

$$\int_{-\pi}^{\pi} \eta(\alpha) d\alpha = 2 \operatorname{Im} \ln[e^{-i\theta/2} \Delta(0)] = \theta. \quad (4.5)$$

Equation (4.3) implies that  $0 \leq (\operatorname{sgn} \theta) \eta(\alpha) \leq 1$  and therefore

$$\int_{-\pi}^{\pi} |\eta(\alpha)| d\alpha = (\operatorname{sgn} \theta) \int_{-\pi}^{\pi} \eta(\alpha) d\alpha = |\theta| \quad (4.6)$$

In the general case,  $T = \sum_{j=1}^{\infty} (e^{i\theta_j} - 1) |g_j\rangle \langle g_j|$  as before, and write  $U_j = U_{j-1}(I + \tau_j |g_j\rangle \langle g_j|)$ , ( $j = 1, 2, \dots$ ). Since  $\tau_j = e^{i\theta_j} - 1$  and since  $(g_j, g_k) = \delta_{jk}$ , it is easy to see that  $U_j$  is unitary and that,

$$\begin{aligned} U_j &= U_0(I + \tau_1 |g_1\rangle \langle g_1|)(I + \tau_2 |g_2\rangle \langle g_2|) \cdots (I + \tau_j |g_j\rangle \langle g_j|) \\ &= U_0 \left( I + \sum_{k=1}^j \tau_k |g_k\rangle \langle g_k| \right) \end{aligned}$$

Therefore  $U_j - U_0 \rightarrow U - U_0$  in  $\mathcal{B}_1$ -norm as  $j \rightarrow \infty$ . As in the last paragraph, let  $\eta_j$  be the real  $L^1[-\pi, \pi]$ -function for the pair  $\{U_j, U_{j-1}\}$  with  $U_j - U_{j-1} = \tau_j |g_j\rangle \langle g_j|$ , an operator of rank one. Then we have from (4.4)–(4.6):

$$\begin{aligned} \operatorname{Im} \Delta_j(\omega) &= e^{i\theta_j/2} + \frac{i}{2} \int_{-\pi}^{\pi} \frac{e^{i\alpha} + \omega}{e^{i\alpha} - \omega} \eta_j(\alpha) d\alpha \\ 0 &\leq (\operatorname{sgn} \theta_j) \eta_j(\alpha) \leq 1, \quad \int_{-\pi}^{\pi} \eta_j(\alpha) d\alpha = \theta_j, \\ \int_{-\pi}^{\pi} |\eta_j(\alpha)| d\alpha &= |\theta_j|, \end{aligned} \quad (4.7)$$

and  $\Delta_j(\omega) = \det[(U_j - \omega)(U_{j-1} - \omega)^{-1}]$ .

Let  $\eta(\alpha) = \sum_{j=1}^{\infty} \eta_j(\alpha)$ . It follows easily from (4.7) that the series converges in  $L^1$ -norm and defines an  $L^1$ -function  $\eta$  and we have

$$\begin{aligned} \int_{-\pi}^{\pi} |\eta(\alpha)| d\alpha &\leq \sum_{j=1}^{\infty} \int_{-\pi}^{\pi} |\eta_j(\alpha)| d\alpha = \sum_{j=1}^{\infty} |\theta_j| \leq \frac{\pi}{2} \|T\|_1, \\ &= \frac{\pi}{2} \|U - U_0\|_1 = \pi \|R_i - R_i^0\|_1, \\ \int_{-\pi}^{\pi} \eta(\alpha) d\alpha &= \sum_{j=1}^{\infty} \theta_j = -i \ln \det(I + T). \end{aligned} \quad (4.8)$$

In theorem A.5 (ii) we have that

$$\begin{aligned} \ln \det[(U_n - \omega)(U_0 - \omega)^{-1}] &= \ln \det[I + (U_n - U_0)(U_0 - \omega)^{-1}] \\ &= \sum_{j=1}^n \ln \Delta_j(\omega) \\ &= \frac{i}{2} \sum_{j=1}^n \theta_j + \frac{i}{2} \int_{-\pi}^{\pi} \frac{e^{i\alpha} + \omega}{e^{i\alpha} - \omega} \left( \sum_{j=1}^n \eta_j(\alpha) \right) d\alpha. \end{aligned} \quad (4.9)$$

In fact the  $U_n - U_0 \rightarrow U - U_0$  in  $\mathcal{B}_1$  and the continuity of the determinant with respect to its argument in  $\mathcal{B}_1$ -norm (theorem A.3 (ii)) lead to

$$\ln \Delta(\omega) = -\frac{i}{2} \sum_{j=1}^{\infty} \theta_j + \frac{i}{2} \int_{-\pi}^{\pi} \frac{e^{i\alpha} + \omega}{e^{i\alpha} - \omega} \eta(\alpha) d\alpha \quad (4.10)$$

and hence

$$\eta(\alpha) = \frac{1}{\pi} \lim_{\rho \uparrow 1} \operatorname{Im} \ln \left[ \exp \left( -\frac{i}{2} \sum_{j=1}^{\infty} \theta_j \right) \Delta(\rho e^{i\alpha}) \right]. \quad (4.11)$$

The transformation  $e^{i\alpha} = \frac{\lambda + i}{\lambda - i}$  or conversely  $\alpha = 2 \cot^{-1} \lambda$  implies that as  $\lambda$  increases from  $-\infty$  to 0 and then to  $+\infty$ ,  $\alpha$  moves from 0 to  $-\pi$  and then from  $\pi$  to 0. If we now define  $\xi: \mathbb{R} \rightarrow \mathbb{R}$  by setting  $\xi(\lambda) = \eta(\alpha) \equiv \eta(2 \cot^{-1} \lambda)$ , then since  $\frac{d\alpha}{d\lambda} = -\frac{2}{1 + \lambda^2}$  one has by (4.8)

$$\int_{-\infty}^{\infty} |\xi(\lambda)| (1 + \lambda^2)^{-1} d\lambda = \frac{1}{2} \int_{-\pi}^{\pi} |\eta(\alpha)| d\alpha \leq \frac{\pi}{4} \|T\|_1$$

and

$$\int_{-\infty}^{\infty} \xi(\lambda) (1 + \lambda^2)^{-1} d\lambda = \frac{1}{2} \int_{-\pi}^{\pi} \eta(\alpha) d\alpha = -\frac{i}{2} \ln \det(I + T)$$

which proves (i) and (ii). The part (iii) is the statement (4.11) which also shows that  $\xi$  is real.

The map  $z \rightarrow \omega = \frac{z + i}{z - i}$  maps the open lower half plane onto the open unit disc and we have  $(U - \omega)^{-1} = \frac{i}{2}(z - i)[1 + (z - i)R_z]$ ,  $(U_0 - \omega)^{-1} = \frac{i}{2}(z - i)[1 + (z - i)R_z^0]$  for  $\operatorname{Im} z < 0$ . Thus by theorem A.5 (ii) and (4.10)

$$\begin{aligned} \operatorname{Tr}[(U - \omega)^{-1} - (U_0 - \omega)^{-1}] &= \frac{i}{2}(z - i)^2 \operatorname{Tr}(R_z - R_z^0) \\ &= -\frac{d}{d\omega} \ln \Delta(\omega) = -\frac{i}{2} \frac{d}{d\omega} \int_{-\pi}^{\pi} \frac{e^{i\alpha} + \omega}{e^{i\alpha} - \omega} \eta(\alpha) d\alpha \\ &= -i \int_{-\pi}^{\pi} \frac{e^{i\alpha}}{(e^{i\alpha} - \omega)^2} \eta(\alpha) d\alpha, \end{aligned}$$

where the interchange of differentiation and integration can easily be justified by noting that the last integral is uniformly convergent for all  $\omega$  such that  $0 \leq |\omega| \leq \delta < 1$ . Therefore for  $\operatorname{Im} z < 0$ ,  $\frac{i}{2}(z - i)^2 \operatorname{Tr}(R_z - R_z^0) = -i \left\{ -\int_{-\infty}^{\infty} \xi(\lambda) \frac{d\alpha}{d\lambda} \right.$   
 $\left. \alpha \lambda \left\{ \left( \frac{\lambda + i}{\lambda - i} \right) \left[ 2i \left( \frac{\lambda - z}{(\lambda - i)(z - i)} \right) \right]^{-2} \right\} = \frac{-i(z - i)^2}{2} \int_{-\infty}^{\infty} \frac{\xi(\lambda)}{(\lambda - z)^2} d\lambda \right.$  which leads to (iii)  
 for  $\operatorname{Im} z < 0$ . Since  $\xi$  is real-valued, an identical formula for  $\operatorname{Im} z > 0$  is obtained by complex conjugation of the one for  $\operatorname{Im} z < 0$ .

Finally, if  $\xi$  and  $\xi'$  are two shift functions satisfying (i) and (ii), then setting  $\zeta(\lambda) = \xi(\lambda) - \xi'(\lambda)$  we have that for  $z = \mu + i\varepsilon$  ( $\varepsilon > 0$ ),

$$\int \frac{\zeta(\lambda) d\lambda}{(\lambda - z)^2} = \int \frac{\zeta(\lambda) d\lambda}{(\lambda - \mu)^2 - \varepsilon^2 - 2i\varepsilon(\lambda - \mu)} = 0.$$

Thus

$$\begin{aligned} 0 &= \operatorname{Im} \int \frac{\zeta(\lambda) d\lambda}{(\lambda - \mu)^2 - \varepsilon^2 - 2i\varepsilon(\lambda - \mu)} = \int \frac{\zeta(\lambda) 2\varepsilon(\lambda - \mu)}{((\lambda - \mu)^2 + \varepsilon^2)^2} d\lambda \\ &= \frac{d}{d\mu} \int \frac{\zeta(\lambda) \varepsilon}{(\lambda - \mu)^2 + \varepsilon^2} d\lambda \end{aligned}$$

or  $\int \zeta(\lambda) \varepsilon ((\lambda - \mu)^2 + \varepsilon^2)^{-1} d\lambda = \text{constant}$  for all  $\mu \in \mathbb{R}$ ,  $\varepsilon > 0$ . By taking limit  $\varepsilon \rightarrow 0+$  of the above expression by theorem 13 of [27],  $\zeta(\mu) = \text{constant}$  almost everywhere which implies uniqueness of  $\xi$  up to an additive constant. ■

Next we consider functions  $\psi$  of  $H$  and  $H_0$  and obtain the trace formula for  $\psi(H) - \psi(H_0)$ .

**Theorem 4.2.** Let  $R_i - R_i^0 \in \mathcal{B}_1$ , and let  $\psi$  be a bounded  $C^1$ -function on  $\mathbb{R}$  such that  $\sup_{\lambda \in \mathbb{R}} |\psi(\lambda)(1 + \lambda^2)| < \infty$  and  $\sup_{\lambda \in \mathbb{R}} |\psi'(\lambda)(1 + \lambda^2)| < \infty$ . Define  $\psi_\varepsilon$  as in (2.9). Then

- (i)  $\psi_\varepsilon(H) - \psi_\varepsilon(H_0) \in \mathcal{B}_1$  for each  $\varepsilon > 0$ .
- (ii) If  $\psi_\varepsilon(H) - \psi_\varepsilon(H_0)$  converges to  $\psi(H) - \psi(H_0)$  in  $\mathcal{B}_1$ -norm, then

$$\operatorname{Tr}[\psi(H) - \psi(H_0)] = \int \psi'(\lambda) \xi(\lambda) d\lambda,$$

where  $\xi$  is the spectral shift function obtained in theorem 4.1.

*Proof.* Let  $z = \mu + i\varepsilon$  ( $0 < \varepsilon < 1$ ) and write  $\Psi_z = (H - i)(H - z)^{-1}$ ,  $\Psi_z^0 = (H_0 - i)(H_0 - z)^{-1}$ . Both  $\Psi_z$  and  $\Psi_z^0$  are bounded and

$$\begin{aligned} R_z - R_z^0 &= R_z[1 + (z - i)R_z^0] - [1 + (z - i)R_z]R_z^0 \\ &= R_z(H_0 - i)R_z^0 - R_z(H - i)R_z^0 \\ &= \Psi_z(R_i - R_i^0)\Psi_z^0, \end{aligned} \tag{4.12}$$

Now let  $R_i - R_i^0 = \sum_{j=1}^{\infty} \gamma_j |f_j\rangle \langle g_j|$  with  $\sum_{j=1}^{\infty} |\gamma_j| = \|R_i - R_i^0\|_1 < \infty$ ,  $\|f_j\| = \|g_j\| = 1$ .

Then by (4.12), Schwarz inequality and lemma 2.2 (ii) we have that

$$\begin{aligned} &\int d\mu (1 + \mu^2)^{-1} \|R_{\mu + i\varepsilon} - R_{\mu + i\varepsilon}^0\|_1 \\ &\leq \sum_{j=1}^{\infty} |\gamma_j| \int d\mu (1 + \mu^2)^{-1} \|\Psi_{\mu + i\varepsilon} f_j\| \|\Psi_{\mu - i\varepsilon}^0 g_j\| \\ &\leq \|R_i - R_i^0\|_1 2\pi(1 + \varepsilon^{-1}). \end{aligned}$$

Next by the definition of  $\psi_\varepsilon$  and functional calculus,

$$\|\psi_\varepsilon(H) - \psi_\varepsilon(H_0)\|_1 \leq \frac{1}{\pi} \int d\mu |\psi(\mu)| \|\operatorname{Im}(R_{\mu + i\varepsilon} - R_{\mu + i\varepsilon}^0)\|_1$$

$$\begin{aligned}
&\leq \frac{1}{\pi} \left( \sup_{\lambda \in \mathbb{R}} |\psi(\lambda)(1+\lambda^2)| \right) \int d\mu (1+\mu^2)^{-1} \|R_{\mu+i\varepsilon} - R_{\mu+i\varepsilon}^0\|_1 \\
&\leq 2(1+\varepsilon^{-1}) \left( \sup_{\lambda \in \mathbb{R}} |\psi(\lambda)(1+\lambda^2)| \right) \|R_i - R_i^0\|_1
\end{aligned}$$

and this proves (i).

$$\begin{aligned}
\text{Tr}[\psi_\varepsilon(H) - \psi_\varepsilon(H_0)] &= \frac{1}{\pi} \int \psi(\lambda) \text{Im Tr}(R_{\lambda+i\varepsilon} - R_{\lambda+i\varepsilon}^0) d\lambda \\
&= -\frac{\varepsilon}{\pi} \int d\lambda \psi(\lambda) \text{Im} \int \frac{\xi(\mu) d\mu}{(\mu - \lambda - i\varepsilon)^2} \\
&= -\frac{\varepsilon}{\pi} \int \psi(\lambda) \frac{d}{d\lambda} \left( \int \frac{\xi(\mu) d\mu}{(\mu - \lambda)^2 + \varepsilon^2} \right) d\lambda \\
&= -\frac{\varepsilon}{\pi} \psi(\lambda) \left( \int \frac{\xi(\mu) d\mu}{(\mu - \lambda)^2 + \varepsilon^2} \right) \Big|_{-\infty}^{\infty} \\
&\quad + \frac{\varepsilon}{\pi} \int \psi'(\lambda) \left( \int \frac{\xi(\mu)}{(\mu - \lambda)^2 + \varepsilon^2} d\mu \right) d\lambda. \tag{4.13}
\end{aligned}$$

Since

$$\psi(\lambda)(1+\lambda^2) \leq C, \quad (1+\mu^2) = 1 + (\mu - \lambda + \lambda)^2 \leq 2[(1+\lambda^2) + (\mu - \lambda)^2]$$

it follows that the boundary term in (4.13) can be estimated by  $2C\varepsilon/\pi \int \frac{\xi(\mu)}{1+\mu^2} \frac{(1+\lambda^2) + (\mu - \lambda)^2}{(1+\lambda^2)[(\mu - \lambda)^2 + \varepsilon^2]} d\mu$ . Furthermore the integrand in the last expression converges to 0 as  $\lambda \rightarrow \pm \infty$  and is bounded by  $(1+\varepsilon^{-2})|\xi(\mu)|(1+\mu^2)^{-1}$  which is integrable, and hence the boundary term in (4.13) vanishes by an application of dominated convergence theorem. The same estimate allows us to interchange the order of integration in the second term in (4.13) to get

$$\begin{aligned}
\text{Tr}[\psi_\varepsilon(H) - \psi_\varepsilon(H_0)] &= \int \frac{\xi(\mu)}{1+\mu^2} \left[ \frac{\varepsilon}{\pi} \int \frac{\psi'(\lambda)(1+\mu^2)}{(\mu - \lambda)^2 + \varepsilon^2} d\lambda \right] d\mu \\
&= \int \frac{\xi(\mu)}{1+\mu^2} \Psi_\varepsilon(\mu) d\mu. \tag{4.14}
\end{aligned}$$

By theorem 13 of [27] we know that since  $\psi'$  is integrable by hypothesis  $\Psi_\varepsilon(\mu)$  converges to  $(1+\mu^2)\psi'(\mu)$  as  $\varepsilon \rightarrow 0+$ . On the other hand since  $|\psi'(\lambda)(1+\lambda^2)| \leq C'$ , as before we get the estimate

$$\begin{aligned}
|\Psi_\varepsilon(\mu)| &\leq C' \left\{ \frac{\varepsilon}{\pi} \int \frac{d\lambda}{(\lambda - \mu)^2 + \varepsilon^2} + \frac{\varepsilon}{\pi} \int \frac{d\lambda}{1+\lambda^2} \right\} \\
&= C'(1+\varepsilon) \leq 2C',
\end{aligned}$$

and we have the result by dominated convergence theorem

many of the main theorems of this section. For this we define the modified Krein Class of complex-valued functions on  $\mathbb{R}$

$$\tilde{\mathcal{K}} = \{\psi: \mathbb{R} \rightarrow \mathbb{C}: \varphi(\lambda) \equiv (1 + \lambda^2)\psi(\lambda) \in \mathcal{K}\} \quad (4.15)$$

where  $\mathcal{K}$  is the Krein class defined in (2.12).

**Theorem 4.3.** *Let  $R_i - R_i^0 \in \mathcal{B}_1$  and  $D(H) = D(H_0)$ . Then for  $\psi \in \tilde{\mathcal{K}}$ ,  $\psi(H) - \psi(H_0) \in \mathcal{B}_1$  and*

$$\text{Tr}[\psi(H) - \psi(H_0)] = \int \psi'(\lambda) \xi(\lambda) d\lambda.$$

We need a lemma.

**Lemma 4.4.** *Assume the hypotheses of theorem 4.3 and set  $\psi^{(t)}(\lambda) = (\lambda^2 + 1)^{-1}(e^{it\lambda} - 1)$  for  $\lambda \in \mathbb{R}$  and  $t \neq 0$ . Then (i)  $\psi^{(t)}$  is a  $C^1$ -function for  $t \neq 0$  and  $\psi^{(t)}(\lambda)(1 + \lambda^2)$  and  $\psi^{(t)'}(\lambda)(1 + \lambda^2)$  are both bounded in  $\lambda$ . Furthermore,  $|\psi^{(t)'}(\lambda)(1 + \lambda^2)|/|t|^{-1} \leq 3$ .*

$$(ii) \quad \psi_\varepsilon^{(t)}(\lambda) = \mp \chi_\pm(t) \left[ \frac{1 - e^{i\lambda t} e^{-\varepsilon|\lambda|}}{1 + (\lambda \pm i\varepsilon)^2} \mp \varepsilon \frac{1 - e^{-|\lambda|}}{(\lambda - i)^2 + \varepsilon^2} \right] \text{ for } t > 0 \text{ and } t < 0 \text{ respectively}$$

and  $\varepsilon > 0$ , where  $\chi_\pm$  are the indicator functions defined at the end of § 3.

$$(iii) \quad \psi^{(t)}(H) - \psi^{(t)}(H_0) \in \mathcal{B}_1 \text{ and}$$

$$\|\psi^{(t)}(H) - \psi^{(t)}(H_0)\|_1 \leq (2\|R_i - R_i^0\|_1 + \|R_{-i} - R_{-i}^0\|_1)|t|.$$

(iv)  $\psi_\varepsilon^{(t)}(H) - \psi_\varepsilon^{(t)}(H_0) \in \mathcal{B}_1$  for every  $\varepsilon > 0$ ,  $t \neq 0$  and  $\psi_\varepsilon^{(t)}(H) - \psi_\varepsilon^{(t)}(H_0)$  converges to  $\psi^{(t)}(H) - \psi^{(t)}(H_0)$  in  $\mathcal{B}_1$ -norm as  $\varepsilon \rightarrow 0+$ .

$$(v) \quad \text{Tr}[\psi^{(t)}(H) - \psi^{(t)}(H_0)] = \int \psi^{(t)'}(\lambda) \xi(\lambda) d\lambda.$$

*Proof.* The part (i) follows by direct verification. For example,  $\frac{d\psi^{(t)}}{d\lambda}(\lambda) = \frac{ite^{it\lambda}}{1 + \lambda^2} - \frac{(e^{it\lambda} - 1)2\lambda}{(\lambda^2 + 1)^2}$  so that

$$|\psi^{(t)'}(\lambda)(1 + \lambda^2)| = |ite^{it\lambda} - \frac{2\lambda}{1 + \lambda^2}(e^{it\lambda} - 1)| \leq 3|t|.$$

To evaluate  $\psi_\varepsilon^{(t)}(\lambda) \equiv \frac{\varepsilon}{\pi} \int \frac{\psi^{(t)}(\mu)}{(\mu - \lambda)^2 + \varepsilon^2} d\mu = \frac{\varepsilon}{\pi} \int \frac{e^{it\mu} - 1}{(1 + \mu^2)((\mu - \lambda)^2 + \varepsilon^2)} d\mu$ , we employ

the method of complex integration by taking a semi-circular contour of radius  $R$  about origin in the upper half complex plane (for  $t > 0$ ). It is easy to see that the contribution to the integral from the semi-circular arc converges to 0 as  $R \rightarrow \infty$  and what remains are the residues at the two enclosed simple poles viz, at  $z = i$  and  $z = \lambda + i\varepsilon$ . This leads to

$$\psi_\varepsilon^{(t)}(\lambda) = \frac{e^{i\lambda t - \varepsilon t} - 1}{(\lambda + i\varepsilon)^2 + 1} + \varepsilon \frac{e^{-t} - 1}{(\lambda - i)^2 + \varepsilon^2} \quad \text{for } t > 0.$$

For  $t < 0$ , one has to take a semi-circular contour in the lower half plane and these two together lead to (ii).

By functional calculus, one writes

$$\begin{aligned}\psi^{(t)}(H) - \psi^{(t)}(H_0) &= R_i(e^{itH} - I)R_{-i} - R_i^0(e^{itH_0} - I)R_{-i}^0 \\ &= (R_i - R_i^0)(e^{itH} - I)R_{-i} + R_i^0(e^{itH_0} - I)(R_{-i} - R_{-i}^0) \\ &\quad + R_i^0(e^{itH} - e^{itH_0})R_i\{(H - i)R_{-i}\}.\end{aligned}\quad (4.16)$$

The first term in (4.16) can be written as  $(R_i - R_i^0) \int_0^t e^{isH} (iHR_{-i}) ds$  and hence is in  $\mathcal{B}_1$  and admits a trace-norm estimate  $|t| \|R_i - R_i^0\|_1$ . An identical consideration leads to a trace-norm estimate for the second term in (4.16) by  $|t| \|R_{-i} - R_{-i}^0\|_1$ . Since  $D(H) = D(H_0)$ , the third term in (4.16) can be written as in (2.6)

$$i \int_0^t e^{i(t-s)H_0} R_i^0 V R_i e^{isH} \{(H - i)R_{-i}\} ds.$$

Since  $R_i^0 V R_i = R_i^0 - R_i \in \mathcal{B}_1$  and since  $\|(H - i)R_{-i}\| = 1$ , by a reasoning similar to the discussion following the proof of theorem 2.1, we conclude that the above integral exists in  $\mathcal{B}_1$ -norm and its trace-norm can be estimated by  $\|R_i - R_i^0\|_1 |t|$ . This proves (iii).

We prove (iv) for  $t > 0$ , the case for  $t < 0$  being similar. That  $\psi_\varepsilon^{(t)}(H) - \psi_\varepsilon^{(t)}(H_0) \in \mathcal{B}_1$  follows from (i) and theorem 4.2. By (ii) and functional calculus,

$$\begin{aligned}[\psi_\varepsilon^{(t)}(H) - \psi_\varepsilon^{(t)}(H_0)] - [\psi^{(t)}(H) - \psi^{(t)}(H_0)] \\ &= \varepsilon(1 - e^{-t})[R_{i+i\varepsilon}R_{i-i\varepsilon} - R_{i+i\varepsilon}^0R_{i-i\varepsilon}^0] \\ &\quad - [\{R_{i-i\varepsilon}R_{-i-i\varepsilon} - R_{i-i\varepsilon}^0R_{-i-i\varepsilon}^0\} - \{R_iR_{-i} - R_i^0R_{-i}^0\}] \\ &\quad + [e^{-t}\{R_{i-i\varepsilon}e^{itH}R_{-i-i\varepsilon} - R_{i-i\varepsilon}^0e^{itH_0}R_{-i-i\varepsilon}^0\} \\ &\quad - \{R_i e^{itH}R_{-i} - R_i^0 e^{itH_0}R_{-i}^0\}].\end{aligned}\quad (4.17)$$

Now by the first and second resolvent identity (since  $D(H) = D(H_0)$ ), we have

$$\begin{aligned}\varepsilon[R_{i+i\varepsilon}R_{i-i\varepsilon} - R_{i+i\varepsilon}^0R_{i-i\varepsilon}^0] \\ &= \frac{1}{2i}\{(R_{i+i\varepsilon} - R_{i+i\varepsilon}^0) - (R_{i-i\varepsilon} - R_{i-i\varepsilon}^0)\} \\ &= \frac{1}{2i}[\{R_{i+i\varepsilon}(H - i)\}(R_i - R_i^0)\{(H_0 - i)R_{i+i\varepsilon}^0\} \\ &\quad - \{R_{i-i\varepsilon}(H - i)\}(R_i - R_i^0)\{(H_0 - i)R_{i+i\varepsilon}^0\}]\end{aligned}$$

$\rightarrow 0$  in  $\mathcal{B}_1$ -norm as  $\varepsilon \rightarrow 0$  since  $(H + i)R_{-i \pm i\varepsilon}$  and  $(H_0 - i)R_{i \pm i\varepsilon}^0 \rightarrow I$  in operator norm. This shows that the first term in (4.17) converges to 0 in  $\mathcal{B}_1$  and the second term in (4.17) can similarly be shown to converge to 0 in  $\mathcal{B}_1$ . For the third term in (4.17) we again use the two resolvent identities to see that the result (iv) follows if

$$\|(R_{i \pm i\varepsilon}e^{itH} - R_{i \pm i\varepsilon}^0e^{itH_0}) - (R_i e^{itH} - R_i^0 e^{itH_0})\|_1 \rightarrow 0$$



$\rightarrow 0+$ . But the above expression

$$\begin{aligned} &= \|\pm i\varepsilon [R_{i\pm i\varepsilon} e^{itH} R_i - R_{i\pm i\varepsilon}^0 e^{itH_0} R_i^0]\|_1 \\ &= \varepsilon \|R_{i\pm i\varepsilon} e^{itH} (R_i - R_i^0) + \{R_{i\pm i\varepsilon} (H - i)\} R_i (e^{itH} - e^{itH_0}) R_i^0 \\ &\quad + [\{R_{i\pm i\varepsilon} (H - i)\} (R_i - R_i^0) e^{itH_0} R_{i\pm i\varepsilon}^0]\|_1. \end{aligned} \quad (4.18)$$

Since  $\|R_{i\pm i\varepsilon}\|$  and  $\|(H - i)R_{i\pm i\varepsilon}\|$  are bounded in  $\varepsilon$  for  $\varepsilon$  sufficiently small, the contributions from the first and third terms are zero in the limit  $\varepsilon \rightarrow 0$ . For the second term in (4.18) we need only to observe from (2.6) that

$$\begin{aligned} R_i (e^{itH} - e^{itH_0}) R_i^0 &= i \int_0^t e^{i(t-s)H} R_i V R_i^0 e^{isH_0} ds = \\ &= -i \int_0^t e^{i(t-s)H} (R_i - R_i^0) e^{isH_0} ds. \end{aligned}$$

So part (v) follows from (iii), (iv) and theorem 4.2 (ii).  $\blacksquare$

Proof of theorem 4.3. By (2.12) and (4.15),

$$\begin{aligned} \psi(\lambda) &= (\lambda^2 + 1)^{-1} \left[ \int \frac{e^{it\lambda} - 1}{it} v(dt) + C \right] \\ &= \int \frac{\psi^{(t)}(\lambda)}{it} v(dt) + C(\lambda^2 + 1)^{-1}. \end{aligned}$$

It is clear that  $(H^2 + I)^{-1} - (H_0^2 + I)^{-1} \in \mathcal{B}_1$ . Also by lemma 4.4 (iii)

$$\begin{aligned} &\left\| \int \frac{\psi^{(t)}(H) - \psi^{(t)}(H_0)}{it} v(dt) \right\|_1 \\ &\leq |v|(\mathbb{R}) (2\|R_i - R_i^0\|_1 + \|R_{-i} - R_{-i}^0\|_1), \end{aligned}$$

hence  $\psi(H) - \psi(H_0) \in \mathcal{B}_1$ . By theorem 4.1 (iv),

$$\begin{aligned} \text{Tr}[(H^2 + I)^{-1} - (H_0^2 + I)^{-1}] &= \text{Im Tr}[(H - i)^{-1} - (H_0 - i)^{-1}] \\ &= -\text{Im} \int \frac{\xi(\lambda)}{(\lambda - i)^2} d\lambda \\ &= -\int \frac{2\lambda}{(\lambda^2 + 1)^2} \xi(\lambda) d\lambda = \int (d/d\lambda) \{(\lambda^2 + 1)^{-1}\} \xi(\lambda) d\lambda. \end{aligned}$$

On the other hand by lemma 4.4 (v),

$$\begin{aligned} \mathcal{J} &\equiv \text{Tr} \int \frac{\psi^{(t)}(H) - \psi^{(t)}(H_0)}{it} v(dt) \\ &= \int \frac{v(dt)}{it} \text{Tr}[\psi^{(t)}(H) - \psi^{(t)}(H_0)] \\ &= \int v(dt) \int \frac{\psi^{(t)}(\lambda) - \psi^{(t)}(\lambda_0)}{it} v(d\lambda) \end{aligned}$$

Since by theorem 4.1 (i),  $\xi(\lambda)(1 + \lambda^2)^{-1} \in L^1$  and by lemma 4.4 (i),  $|\psi^{(n)}(\lambda)(1 + \lambda^2)| \leq C|t|$ , and since  $\nu$  is a finite measure, we can interchange the order of integration in the above and get

$$\begin{aligned} \mathcal{S} &= \int \xi(\lambda) d\lambda \int \frac{1}{it} \left\{ \frac{ite^{it\lambda}}{\lambda^2 + 1} - \frac{2\lambda(e^{it\lambda} - 1)}{(\lambda^2 + 1)^2} \right\} \nu(dt) \\ &= \int \xi(\lambda) d\lambda \frac{d}{d\lambda} \left[ \int \frac{e^{it\lambda} - 1}{it(\lambda^2 + 1)} \nu(dt) \right]. \end{aligned}$$

The interchange of differentiation and integration in the last step is justified since

$$\left| \frac{1}{it} \left\{ \frac{ite^{it\lambda}}{\lambda^2 + 1} - \frac{2\lambda(e^{it\lambda} - 1)}{(\lambda^2 + 1)^2} \right\} \right| \leq 3(\lambda^2 + 1)^{-1}$$

for all  $t \neq 0$  and hence the concerned integral converges uniformly. This completes the proof. ■

*Remark 4.5(i).* If  $\psi \in \mathcal{S}(\mathbb{R})$ , the Schwartz class of smooth function of rapid decrease at  $\infty$ , then so is  $\varphi(\lambda) = \psi(\lambda)(1 + \lambda^2)$  and thus by remark 2.7 (i)  $\mathcal{S}(\mathbb{R}) \subseteq \tilde{\mathcal{K}}$ . Also all functions of the type  $(\lambda - z)^{-m}$  (for integer  $m \geq 1$ ) are in  $\tilde{\mathcal{K}}$ . Krein in his original work considered functions  $\psi$  admitting integral representation

$$\psi(\lambda) = \int (\lambda - z)^{-1} d\mu(z), \quad (4.19)$$

where  $\mu$  is complex measure on the set of non-real points in  $\mathbb{C}$  satisfying for  $z = x + iy$ ,  $\int |y|^{-j} |d\mu(z)| < \infty$  ( $j = 1, 2$ ). A simple calculation as in the proof of theorem 3.3 (v) shows that since  $\int \exp(-|ty|) |d\mu(z)| < \infty$  for every  $t \neq 0$ , this class of functions are contained in  $\tilde{\mathcal{K}}$ .

(ii) Let  $J$  be a real open interval in  $\rho(H) \cap \rho(H_0)$ , and let  $\psi \in C_0^\infty(J)$ , the class of smooth functions with compact support in  $J$ . Then by functional calculus  $\psi(H) = \psi(H_0) = 0$  and hence by the trace formula in theorem 4.3,

$$\int \xi(\lambda) \psi'(\lambda) d\lambda = 0 \quad \forall \quad \psi \in C_0^\infty(J). \quad (4.20)$$

Since  $\xi(\lambda)(1 + \lambda^2)^{-1} \in L^1$ , it follows that  $\xi \in L_{\text{loc}}^1(\mathbb{R})$  and hence as in remark 2.7 (iii), the equation (4.20) can be viewed as  $\langle \xi', \psi \rangle = 0$  for  $\psi \in C_0^\infty(J)$  where  $\xi'$  is the distributional derivative and we conclude that  $\xi$  is a constant in  $J$ .

Thus if  $H$  and  $H_0$  are bounded below, which happens for many Schrödinger operators ([1], [23]), the shift function is constant in the neighbourhood of  $-\infty$  and can be chosen to be zero there.

## 5. Applications

First of all we want to mention the relation between the spectral shift function and scattering theory. One of the earliest results in this direction is due to Birman and Krein [7]. More details can be found in ([20], [5]).

Let  $H$  and  $H_0$  be self-adjoint and let  $R_z - R_z^0 \in \mathcal{B}_1$  for some  $z \in \rho(H) \cap \rho(H_0)$ . Define the operators  $\Omega_{\pm}$  (if they exist) as:

$$\Omega_{\pm}(H, H_0) \equiv \Omega_{\pm} = s - \lim_{t \rightarrow \pm \infty} e^{iHt} e^{-iH_0 t} E_{ac}^0, \quad (5.1)$$

where  $E_{ac}^0$  is the projection onto the absolutely continuous subspace of  $H_0$ . If  $\Omega_{\pm}$  exist, then they are partial isometries with initial set  $E_{ac}^0 \mathcal{H}$  and final set closed subspaces of  $E_{ac} \mathcal{H}$ , and satisfy the intertwining property:

$$\Omega_{\pm} H_{0,ac} = H_{ac} \Omega_{\pm}, \quad (5.2)$$

where  $H_{ac}$  and  $H_{0,ac}$  are the absolutely continuous parts of  $H$  and  $H_0$  respectively,  $E_{ac}$  is the projection onto the absolutely continuous subspace of  $\mathcal{H}$ . The wave operators are said to be *complete* if their final sets are both  $E_{ac} \mathcal{H}$  i.e. if

$$\text{Range } \Omega_{+} = \text{Range } \Omega_{-} = E_{ac} \mathcal{H}. \quad (5.3)$$

If  $\Omega_{\pm}$  exist and are complete, then one defines the *scattering operator*  $S$ :

$$S = \Omega_{+}^{*} \Omega_{-}, \quad (5.4)$$

where one observes that  $S$  commutes with  $H_0$  and  $SS^{*} = S^{*}S = E_{ac}^0$ . In such a case,  $E_{ac}^0 \mathcal{H}$  admits a direct integral representation (upto unitary isomorphism) [11]:

$$E_{ac}^0 \mathcal{H} = \int^{\oplus} \mathcal{H}_{\lambda} d\lambda,$$

so that

$$H_{0,ac} = \int^{\oplus} \lambda d\lambda \text{ and} \quad (5.5)$$

$$S = \int^{\oplus} S(\lambda) d\lambda.$$

For almost all  $\lambda$ ,  $S(\lambda)$  is a unitary operator in  $\mathcal{H}_{\lambda}$  and is called the *scattering matrix* or *on-shell scattering operator*. One can also define a self-adjoint operator-valued function  $\Pi(\lambda)$  (called the *phase shift*) such that

$$S(\lambda) = \exp(-2\pi i \Pi(\lambda)). \quad (5.6)$$

We now state a theorem (without proof) which is typical of scattering theory and relates the shift function  $\xi$  with the phase shift operator  $\Pi(\lambda)$  in (5.6).

**Theorem 5.1.** *Let  $R_z - R_z^0 \in \mathcal{B}_1$  for some  $z$  in  $\rho(H) \cap \rho(H_0)$ . Then the wave operators in (5.1) exist and are complete. Furthermore,  $\Pi(\lambda) \in \mathcal{B}_1(\mathcal{H}_{\lambda})$  for almost all  $\lambda$ ,  $\xi(\lambda) = \text{Tr}(\Pi(\lambda))$  so that*

$$\det S(\lambda) = \exp(-2\pi i \xi(\lambda)), \quad (5.7)$$

where  $\xi(\lambda)$  is the spectral shift function obtained in theorem 4.1.

For an introduction to scattering theory and a proof of the above theorem, the reader is referred to [11] and [5].

relation between the average time-delay in a scattering process and the (distributional) derivative of the associated shift function  $\xi(\lambda)$  [15].

In a series of papers the authors of [6] and [13] used the trace formula, in particular (2.17) appearing in remark 2.7 (v), to compute the Witten index in super-symmetric quantum mechanics. A typical theorem which we state without proof, is the following.

**Theorem 5.2.** *Let  $A$  be a closed operator in  $\mathcal{H}$  such that  $\exp(-A^*A) - \exp(-AA^*)$  is trace-class. If we assume that the associated shift function  $\xi(\lambda)$  is right continuous at 0 and if we define the Witten index  $W(A) \equiv \lim_{t \rightarrow \infty} \text{Tr}(e^{-tA^*A} - e^{-tAA^*})$ , then  $W(A) = -\xi(0+)$ .*

From the remark 2.7 (v), it is clear that  $[\exp(-tA^*A) - \exp(-tAA^*)] \in \mathcal{B}_1$  for all  $t > 2$  and

$$\text{Tr}[(\exp(-tA^*A) - \exp(-tAA^*))] = -t \int_0^\infty -e^{-t\lambda} \xi(\lambda) d\lambda.$$

Then the result of theorem 5.2 follows from this expression and the hypothesis of right continuity. Under further assumptions, the authors in [13] prove an invariance property of the index for a class of perturbations.

Another interesting application is in relation to a pair of projections  $P, Q$  in a Hilbert space  $\mathcal{H}$ . An ordered pair of projections  $(P, Q)$  is said to be a *Fredholm pair* if  $G \equiv QP$ :  $\text{Range } P \rightarrow \text{Range } Q$  is Fredholm i.e. if  $\mathcal{R}(G) \equiv \text{Range } G$  is closed and if  $\mathcal{N}(G) (\equiv \text{the null space of } G)$  and  $\mathcal{R}(G)$  have finite dimension and co-dimension respectively. In such a case, we define the index of the pair

$$\text{Ind}(P, Q) \equiv \dim \mathcal{N}(G) - \dim \mathcal{R}(G)^\perp.$$

Then the following theorem ([2], [3]) can be proven.

**Theorem 5.3.** *Set  $\mathcal{H}_{mn}(m, n = 0, 1) \equiv \{f \in \mathcal{H} | Pf = mf, Qf = nf\}$ .*

(i) *If  $(P, Q)$  is a Fredholm pair, then  $m_1 = \dim \mathcal{H}_{10}$  and  $m_{-1} = \dim \mathcal{H}_{01}$  are finite and*

$$\text{Ind}(P, Q) = \dim \mathcal{H}_{10} - \dim \mathcal{H}_{01} \equiv m_1 - m_{-1}.$$

(ii) *If  $A \equiv P - Q \in \mathcal{B}_1$ , then  $(P, Q)$  is a Fredholm pair and  $A^{2n+1} \in \mathcal{B}_1$  for all positive integer  $n$  and  $\text{Tr } A^{2n+1} = \text{Tr } A = \text{Ind}(P, Q) = m_1 - m_{-1}$ , an integer.*

(iii) *If  $A \in \mathcal{B}_1$ , then the perturbation determinant  $\Delta(z)$  (for  $\text{Im } z \neq 0$ ) is given as*

$$\Delta(z) = \left( \frac{z-1}{z} \right)^{m_1 - m_{-1}}.$$

The shift function  $\xi$  in this case is given by

$$\xi(\lambda) = \begin{cases} 0 & \text{if } \lambda \notin [0, 1] \\ m_1 - m_{-1} & \text{if } \lambda \in [0, 1]. \end{cases}$$

These results find application in the study of charge transport phenomenon in

ger Hall effect [4] and a proof of the above theorem and its generalizations can be found in [2]. A recent survey of some further applications of the trace formula can be found in the lecture notes of Simon [24].

## Appendix

In the first part we state some of the standard results on boundary values of functions analytic in half-plane and unit disc. These have been used in §2 and §4. Next we define the perturbation determinant and study some of its properties.

**Theorem A.1.** *Let  $F(z)$  be analytic in the open upper half plane  $\{z: \operatorname{Im} z > 0\}$  with  $\operatorname{Im} F(z) \leq C$  for some constant  $C$ , and  $|F(z)| = O\left(\frac{1}{\operatorname{Im} z}\right)$  as  $\operatorname{Im} z \rightarrow \infty$ . Then there exists a unique real valued  $L^1$  function  $\zeta$  on  $\mathbb{R}$  given by  $\zeta(\lambda) = (1/\pi) \lim_{\varepsilon \rightarrow 0+} \operatorname{Im} F(\lambda + i\varepsilon)$  for almost all  $\lambda$  (Lebesgue) such that*

$$F(z) = \int_{-\infty}^{\infty} \frac{\zeta(\lambda)}{\lambda - z} d\lambda.$$

Such a function  $F$ , analytic in the open upper half plane such that  $\operatorname{Im} F(z) \geq 0$ , is called a Herglotz function. A Herglotz function  $F$  satisfying  $|F(z)| = O\left(\frac{1}{\operatorname{Im} z}\right)$  as  $\operatorname{Im} z \rightarrow \infty$  admits the following representation (see theorem B3 of [29]):  $F(z) = \int_{-\infty}^{\infty} \frac{d\sigma(\lambda)}{\lambda - z}$ , where  $\sigma$  is a right continuous non-decreasing bounded function on  $\mathbb{R}$ . The further restriction  $\operatorname{Im} F(z) \leq C$  leads to the absolute continuity of  $\sigma$  such that  $\zeta(\lambda) (= \sigma'(\lambda) \text{ a.e.})$  is integrable.

**Theorem A.2.** *Let  $G(\omega)$  be analytic in open unit disc  $|\omega| < 1$  and let  $0 \leq \operatorname{Im} G(\omega) \leq C$  for some constant  $C$ . Then there exists a real valued function  $\eta$  in  $L^1[-\pi, \pi]$  such that*

$$G(\omega) = \operatorname{Re} G(0) + \frac{i}{2} \int_{-\pi}^{\pi} \frac{e^{i\alpha} + \omega}{e^{i\alpha} - \omega} \eta(\alpha) d\alpha,$$

$$\eta(\alpha) = (1/\pi) \lim_{\rho \uparrow 1} \operatorname{Im} G(\rho e^{i\alpha}) \text{ for almost all } \alpha.$$

For a proof of this, see for example pages 189–198 of [21].

In analogy with the case of operators in finite dimensional Hilbert space or of trace class operators in infinite dimensional Hilbert space, the determinant  $\det(I + A)$ ,  $A \in \mathcal{B}_1$ , is defined as:

$$\det(I + A) \equiv \prod_{j=1}^{\infty} (1 + \lambda_j(A)), \quad (\text{A.1})$$

where  $\lambda_j(A)$ 's are the eigenvalues of  $A$  counted as many times as their multiplicities.

The above definition makes sense since  $\sum_{j=1}^{\infty} |\lambda_j(A)| \leq \|A\|_1$  for  $A \in \mathcal{B}_1$ . The following properties of determinant can be proven (see [14] for further details).

**Theorem A.3.** Let  $A \in \mathcal{B}_1$ , and let  $\{\lambda_j(A)\}$  be the eigenvalues of  $A$ . Then

- (i)  $\det(I + A) = \exp\left[\int_{\Gamma} dz \operatorname{Tr}\{A(I + zA)^{-1}\}\right]$ , where  $\Gamma$  is a rectifiable path in  $\mathbb{C}$  joining 0 and 1 such that none of the points  $\{-\lambda_j(A)^{-1}\}$  lies on  $\Gamma$ ,
- (ii) given  $\varepsilon > 0$  there exists  $\delta > 0$  such that for every  $B \in \mathcal{B}_1$  with  $\|A - B\|_1 < \delta$ ,  $|\det(I + A) - \det(I + B)| < \varepsilon$ , i.e.  $A \rightarrow \det(I + A)$  is continuous in  $\mathcal{B}_1$ -norm.
- (iii) If  $B \in \mathcal{B}_1$ , then

$$\det[(I + A)(I + B)] = \det(I + A) \cdot \det(I + B).$$

$$(iv) |\det(I + A)| \leq e^{\|A\|_1}.$$

$$(v) \text{ For every unitary operator } S, \det(I + A) = \det(I + SAS^{-1}).$$

*Proof.* Since  $A$  is compact, the intersection of the set  $\{-\lambda_j(A)^{-1}\}$  with any bounded subset of  $\mathbb{C}$  is finite (could be empty). For  $z \in \mathbb{C}$ , define

$$D(z) \equiv \det(I + zA) = \begin{cases} \prod_{j=1}^{\infty} (1 + z\lambda_j(A)) & \text{if } z \neq -\lambda_j(A)^{-1} \text{ for any } j, \\ 0 & \text{if } z = -\lambda_j(A)^{-1} \text{ for some } j. \end{cases}$$

Then  $D(z)$  is analytic in the complex plane with zeros accumulating at infinity. Taking the logarithmic derivative, at points away from these zeros,

$$\frac{D'(z)}{D(z)} = \sum_{j=1}^{\infty} \frac{\lambda_j(A)}{1 + z\lambda_j(A)} = \operatorname{Tr} A(I + zA)^{-1}. \quad (\text{A.2})$$

Let  $\Gamma$  be a rectifiable curve in  $\mathbb{C}$  joining 0 and 1 such that none of  $-\lambda_j(A)^{-1}$ 's lie on  $\Gamma$ . Integrating both the sides over  $\Gamma$  and then taking the exponential we get the required result. A priori it seems, the integral depends on the path. But if we extend  $\Gamma$  to a closed contour by taking another rectifiable curve  $\Gamma'$  (say) from 1 to 0 such that none of the  $-\lambda_j(A)^{-1}$ 's lie on  $\Gamma'$ , then  $\operatorname{Int}(\Gamma \cup \Gamma')$  contains at most finitely many  $-\lambda_j(A)^{-1}$ , say  $-\lambda_1(A)^{-1}, -\lambda_2(A)^{-1}, \dots, -\lambda_k(A)^{-1}$ . Then the integration over  $\Gamma \cup \Gamma'$  has the contribution  $\sum_{j=1}^k 2\pi i m_j$ , where  $m_j$  is the multiplicity of  $\lambda_j(A)$ , which equals identity on exponentiation. This proves (i).

By the resolvent identity

$$(I + zA)^{-1} - (I + zB)^{-1} = z(I + zA)^{-1}(B - A)(I + zB)^{-1}$$

or

$$(I + zB)^{-1} = [I + z(I + zA)^{-1}(B - A)]^{-1}(I + zA)^{-1}.$$

If  $B \in \mathcal{B}_1$  be such that

$$\|B - A\|_1 \leq \min_{z \in \Gamma} \{ |z| \|(I + zA)^{-1}\| \}^{-1}, \quad (\text{A.3})$$

then  $[I + z(I + zA)^{-1}(B - A)]^{-1}$  exists as a Neumann series, and

$$\sup_{z \in \Gamma} \|(I + zB)^{-1}\| \leq C \sup_{z \in \Gamma} \|(I + zA)^{-1}\|, \quad (\text{A.4})$$

re the constant  $C$  depends only on  $\Gamma$  and  $A$ . Let  $L(\Gamma)$  be the length of the arc  $\Gamma$ .  
e

$$\begin{aligned}\|A(I+zA)^{-1} - B(I+zB)^{-1}\|_1 &= \|(I+zB)^{-1}(A-B)(I+zA)^{-1}\|_1 \\ &\leq \|(I+zB)^{-1}\| \|A-B\|_1 \|(I+zA)^{-1}\|,\end{aligned}$$

llows from (A.4) that for any  $\kappa > 0$ , there exists a  $\delta > 0$  such that

$$\begin{aligned}&\left| \int_{\Gamma} \text{Tr} A(I+zA)^{-1} - B(I+zB)^{-1} \} dz \right| \\ &\leq \int_{\Gamma} \|A(I+zA)^{-1} - B(I+zB)^{-1}\|_1 dz \\ &\leq L(\Gamma) C \left\{ \sup_{z \in \Gamma} \|(I+zA)^{-1}\|^2 \|A-B\|_1 \right\} \\ &< \kappa\end{aligned}$$

never  $\|A-B\|_1 < \delta$ . Using the inequality  $|e^z - 1| \leq \sqrt{2}e^{|z|}|z|$ , and choosing  $\kappa$  sufficiently small we get

$$\begin{aligned}&|\det(I+B) - \det(I+A)| \\ &= \left| \exp \left[ \int_{\Gamma} dz \text{Tr} \{B(I+zB)^{-1}\} \right] - \exp \left[ \int_{\Gamma} dz \text{Tr} \{A(I+zA)^{-1}\} \right] \right| \\ &\leq \left| \exp \left[ \int_{\Gamma} dz \text{Tr} \{A(I+zA)^{-1}\} \right] \right| \\ &\quad \left| \exp \left[ \int_{\Gamma} dz \text{Tr} \{B(I+zB)^{-1} - A(I+zA)^{-1}\} \right] - 1 \right| \\ &\leq \sqrt{2}e\kappa |\det(I+A)|\end{aligned}$$

ch can be made arbitrarily small by choosing  $\delta$  appropriately.

et  $\{P_n\}$  be a sequence of finite rank projections such that  $P_n \uparrow I$ . Then  $P_n A P_n$  and  $P_n B P_n$  converges to  $A$  and  $B$  respectively in  $\mathscr{B}_1$ -norm as  $n \rightarrow \infty$ . Hence

$$\begin{aligned}(I + P_n A P_n)(I + P_n B P_n) - (I + A)(I + B) \\ = (P_n A P_n - A) + (P_n B P_n - B) + P_n A P_n B P_n - AB \rightarrow 0\end{aligned}$$

$\mathscr{B}_1$ -norm as  $n \rightarrow \infty$ . Note that

$$\det(I + P_n A P_n) = \det(P_n + P_n A P_n),$$

re the determinant on the right hand side is taken on the finite dimensional  
bert space  $P_n \mathscr{H}$ . By (ii),

$$\det[(I+A)(I+B)]$$

$$\begin{aligned}
&= \lim_{n \rightarrow \infty} \det[(P_n + P_n A P_n)(P_n + P_n B P_n)] \\
&= \lim_{n \rightarrow \infty} \det(P_n + P_n A P_n) \det(P_n + P_n B P_n) \\
&= \lim_{n \rightarrow \infty} \det(I + P_n A P_n) \det(I + P_n B P_n) \\
&= \det(I + A) \det(I + B),
\end{aligned}$$

which proves (iii). By (A.1), and the inequality  $1 + x < e^x$  for  $x > 0$ , we get

$$\begin{aligned}
|\det(I + A)| &\leq \prod_{j=1}^{\infty} (1 + |\lambda_j(A)|) \\
&\leq \exp\left(\sum_{j=1}^{\infty} |\lambda_j(A)|\right) \\
&\leq e^{\|A\|_1}.
\end{aligned}$$

Since  $\sigma(A) = \sigma(SAS^{-1})$  for any unitary operator  $S$ , part (v) follows from (A.1). ■

Next we define the perturbation determinants for two explicit cases and study some of their properties.

*Case I.* Let  $H$  be a self-adjoint operator in  $\mathcal{H}$ . Assume that  $V_1$  and  $V_2$  are two trace class self-adjoint operators, so that  $H_j = H + V_j$  are self-adjoint for  $j = 1, 2$ . For  $\text{Im } z \neq 0$ , define the perturbation determinants

$$\begin{aligned}
\Delta_j(z) &\equiv \det[I + V_j(H - z)^{-1}] \quad \text{for } j = 1, 2, \\
\Delta_{2,1}(z) &\equiv \det[I + (V_2 - V_1)(H_1 - z)^{-1}].
\end{aligned} \tag{A.5}$$

*Case II.* Let  $U, U_1$  and  $U_2$  be three unitary operators in  $\mathcal{H}$  such that  $U_j - U \in \mathcal{B}_1$  for  $j = 1, 2$ . For complex  $\omega$  with  $|\omega| < 1$  define the perturbation determinants

$$\begin{aligned}
\Delta_j(\omega) &\equiv \det[I + (U_j - U)(U - \omega)^{-1}] \quad \text{for } j = 1, 2, \\
\Delta_{2,1} &\equiv \det[(I + (U_2 - U_1)(U_1 - \omega)^{-1})].
\end{aligned} \tag{A.6}$$

We start with the following abstract result.

*Lemma A.4.* Let  $z \rightarrow A(z)$  be a  $\mathcal{B}_1$ -valued analytic function in some domain  $D$  in  $\mathbb{C}$ . Then  $\det(I + A(z))$  is analytic in  $D$ . For all  $z$  for which  $(I + A(z))^{-1} \in \mathcal{B}(\mathcal{H})$ ,  $\ln \det(I + A(z))$  is an analytic function and

$$\frac{d}{dz} \ln \det(I + A(z)) = \text{Tr} \left\{ (I + A(z))^{-1} \frac{dA(z)}{dz} \right\}.$$

*Proof.* As in the proof of theorem A.3, we choose an increasing sequence of finite dimensional projections  $\{P_n\}$  and view  $P_n + P_n A(z) P_n$  as acting in  $P_n \mathcal{H}$ . Then for



ry  $z \in D$ ,

$$\begin{aligned}\Delta(z) &= \det(I + A(z)) \\ &= \lim_{n \rightarrow \infty} \det(I + P_n A(z) P_n) \\ &= \lim_{n \rightarrow \infty} \det(P_n + P_n A(z) P_n) \\ &\equiv \lim_{n \rightarrow \infty} \Delta_n(z).\end{aligned}$$

ce  $A(z)$  is  $\mathcal{B}_1$ -analytic in  $D$ ,  $\Delta_n(z)$  is analytic for each  $n$  and by theorem A.3 (iv),  $|\Delta_n(z)| \leq \exp(\|A(z)\|_1) \leq M$ . This and Cauchy's integral formula implies equicontinuity of  $\{\Delta_n(z)\}$ , and by Ascoli's theorem (relabelling the consequent subsequence) we conclude that  $\Delta_n(z)$  converges to  $\Delta(z)$  as  $n \rightarrow \infty$  uniformly in  $z$  in compact subsets of  $D$  and consequently  $\Delta(z)$  is analytic.

Set  $A_n(z) = P_n A(z) P_n$  and let  $(I + A(z_0))^{-1} \in \mathcal{B}(\mathcal{H})$  for some  $z_0 \in D$ . Then there is an open ball  $\mathcal{U}$  about  $z_0$  such that  $(I + A(z))^{-1} \in \mathcal{B}(\mathcal{H})$  for all  $z \in \mathcal{U}$ . Thus  $\Delta(z) \neq 0$  and hence  $\ln \Delta(z)$  is analytic in  $\mathcal{U}$ . Since  $\Delta_n(z)$  converges to  $\Delta(z)$  uniformly in  $\tilde{\mathcal{U}}$  (a closed ball in  $\mathcal{U}$ ),  $\Delta_n(z) \neq 0$  for  $z \in \tilde{\mathcal{U}}$  and  $n > N$  (depending on  $\tilde{\mathcal{U}}$  only). By the spectral theory of compact operators, we have that  $(I + A_n(z))^{-1} \in \mathcal{B}(\mathcal{H})$  or equivalently  $(I + A_n(z))^{-1} \in \mathcal{B}(P_n \mathcal{H})$  and therefore  $\ln \Delta_n(z)$  is analytic for such  $n$  and  $z$ . For finite dimensional determinants, the formula in this lemma is well known and we have for  $n$  and  $z$  as above,

$$\begin{aligned}\frac{d}{dz} \ln \Delta_n(z) &= \frac{\Delta'_n(z)}{\Delta_n(z)} \\ &= \text{Tr} \left[ (P_n + A_n(z))^{-1} \frac{dA_n(z)}{dz} \right] \\ &= \text{Tr} \left[ (I + A_n(z))^{-1} \frac{dA_n(z)}{dz} \right].\end{aligned}\tag{A.7}$$

For such  $z$  and  $n$  one has the identity:

$$(I + A_n(z))^{-1} - (I + A(z))^{-1} = (I + A_n(z))^{-1} \{A(z) - A_n(z)\} (I + A(z))^{-1}.$$

This implies that for fixed  $z \in \tilde{\mathcal{U}}$ ,  $\|(I + A_n(z))^{-1}\| \leq M(z)$  and  $(I + A_n(z))^{-1}$  converges to  $(I + A(z))^{-1}$  as  $n \rightarrow \infty$  in  $\mathcal{B}_1$ . Since  $\frac{dA(z)}{dz} \in \mathcal{B}_1$ , it follows that  $\frac{dA_n(z)}{dz}$  converges to  $\frac{dA(z)}{dz}$  in  $\mathcal{B}_1$  and hence the right hand side of (A.7) converges pointwise in  $\tilde{\mathcal{U}}$  to

$\frac{d}{dz} \ln \Delta(z)$  in  $\mathcal{B}_1$  and hence the right hand side of (A.7) converges pointwise in  $\tilde{\mathcal{U}}$  to  $\frac{d}{dz} \ln \Delta(z)$ . As for the left hand side of (A.7) we need only to use Cauchy's

$$\frac{d}{dz} \ln \Delta_j(z) = -\text{Tr} \{ (H_j - z)^{-1} - (H - z)^{-1} \} \quad (\text{A.8})$$

for  $j = 1, 2$ .

(ii) The perturbation determinants  $\Delta_{2,1}(\omega), \Delta_j(\omega)$ , ( $j = 1, 2$ ), given by (A.6) are analytic for all complex number  $\omega$  with  $|\omega| < 1$  and have no zeros. Furthermore  $\Delta_{2,1}(\omega)\Delta_1(\omega) = \Delta_2(\omega)$  and

$$\frac{d}{d\omega} \ln \Delta_j(\omega) = -\text{Tr} [(U_j - \omega)^{-1} - (U - \omega)^{-1}] \quad (\text{A.9})$$

for  $j = 1, 2$ .

*Proof.* We shall only prove (i) since the proof of (ii) is identical.

Since for  $\text{Im } z \neq 0$ ,  $(H - z)^{-1}$  is analytic in  $\mathcal{B}(\mathcal{H})$ , so  $V_j(H - z)^{-1}$  is analytic in  $\mathcal{B}_1$ . Hence by lemma A.4,  $\Delta_j(z)$  is analytic in the same domain. Furthermore for  $\text{Im } z \neq 0$  and  $f \in \mathcal{H}$ ,  $[I + V_j(H - z)^{-1}]f = (H_j - z)(H - z)^{-1}f = 0$  implies  $f = 0$  since  $z$  belongs to  $\rho(H) \cap \rho(H_j)$ . Thus,  $\Delta_j(z)$  has no zeros there. Hence by theorem A.4,  $\ln \Delta_j(z)$  is analytic for  $\text{Im } z \neq 0$ , and

$$\begin{aligned} \frac{d}{dz} \ln \Delta_j(z) &= \text{Tr} \left[ [I + V_j(H - z)^{-1}]^{-1} \frac{d}{dz} \{ V_j(H - z)^{-1} \} \right] \\ &= \text{Tr} [ \{ I - V_j(H_j - z)^{-1} \} V_j(H - z)^{-2} ] \\ &= \text{Tr} [ (H - z)^{-1} \{ I - V_j(H_j - z)^{-1} \} V_j(H - z)^{-1} ] \\ &= \text{Tr} [ (H_j - z)^{-1} V_j(H - z)^{-1} ] \\ &= -\text{Tr} \{ (H_j - z)^{-1} - (H - z)^{-1} \}. \end{aligned}$$

Next by theorem A.3 (iii),

$$\begin{aligned} \Delta_{2,1}(z)\Delta_1(z) &= \det[I + (V_2 - V_1)(H_1 - z)^{-1}] \cdot \det[I + V_1(H - z)^{-1}] \\ &= \det[ \{ I + (V_2 - V_1)(H_1 - z)^{-1} \} \{ I + V_1(H - z)^{-1} \} ] \\ &= \det[ I + (V_2 - V_1) \{ (H_1 - z)^{-1} + (H_1 - z)^{-1} V_1(H - z)^{-1} \} + V_1(H - z)^{-1} ] \\ &= \det[ I + (V_2 - V_1)(H - z)^{-1} + V_1(H - z)^{-1} ] \\ &= \det[ I + V_2(H - z)^{-1} ] \\ &= \Delta_2(z). \end{aligned}$$

■

## References

- [1] Amrein W O, Jauch J M and Sinha K B, *Scattering theory in quantum mechanics* (Massachusetts: W A Benjamin) (1977)
- [2] Amrein W O and Sinha K B, On pairs of projections in a Hilbert space to appear in *Linear algebra and its applications* **208/209** (1994) 425–435
- [3] Avron J, Seiler R and Simon B, The index of a pair of projections, *J. Funct. Anal.* **120** (1994) 220–237
- [4] Avron J, Seiler R and Simon B, Charge deficiency, charge transport and comparison of dimensions, *Comm. Math. Phys.* **159**(2) (1994) 399–422
- [5] Baumgartel H and Wollenberg M, *Mathematical scattering theory*, (Berlin: Akademie Verlag) (1983)
- [6] Bolle D, Gesztesy F, Grosse H, Schweiger W and Simon B, Witten index, axial anomaly and Krein's spectral shift function in super-symmetric quantum mechanics, *J. Math. Phys.* **28**(7) (1987) 1512–25
- [7] Birman M S and Krein M G, On the theory of wave and scattering operators, *Soviet Math. Dokl.* **3** (1962) 740–744
- [8] Birman M S and Solomyak M Z, Remarks on the spectral shift function, *Zap. Nauch. Sem. Len. Otdel. Mat. Inst. Steklova, Akad. Nauk. SSSR* **27** (1972) 33–46, (English translation: *J. Sov. Math.* **3** (4) (1975) 408–419)
- [9] Birman M S and Yafaev D R, The spectral shift function. Works by M G Krein and their development (Russian). *Algebra i Analis*, **4** (1992) 1–44
- [10] Clancy K, *Seminormal Operators*, Lecture notes in Mathematics – 742, (Heidelberg: Springer Verlag) (1979)
- [11] Dixmier J, *Von Neumann algebras* (Amsterdam: North Holland) (1981)
- [12] Donoghue (Jr.) W F, *Distributions and Fourier transforms* (New York: Academic Press) (1969)
- [13] Gesztesy F and Simon B, Invariance properties of Witten Index, *J. Funct. Anal.* **79** (1988) 91–102
- [14] Gohberg I C and Krein M G, *Introduction to the theory of linear non-self-adjoint operators* (Translations of Mathematical Monographs, Vol. 18, American Mathematical Society, Providence, R I, 1969)
- [15] Jauch J M, Sinha K B and Misra B N, Time-delay in scattering processes, *Helv. Phys. Acta.* **45** (1972) 398–426
- [16] Kato T, *Perturbation theory for linear operators* (2nd ed.) (New York: Springer Verlag) (1976)
- [17] Kuroda S T, On a generalization of Weinstein-Aronszajn formula and infinite determinant, *Sci. Papers Coll. Gen. Ed. Univ. Tokyo* **11** (1961) 1–12
- [18] Krein M G, On the trace formula in perturbation theory, (Russian) *Math. Sb.* **33** (1953) 597–626
- [19] Krein M G, On perturbation determinants and a trace formula for unitary and self-adjoint operators, *Soviet Math. Dokl.* **3** (1962) 707–710
- [20] Krein M G, On certain new studies in the perturbation theory for self-adjoint operators, (107–172), in *Topics in Differential and Integral equations, and operator theory* (Ed. I Gohberg), OT **7** (Basel: Birkhauser-Verlag) (1983)
- [21] Nevanlinna R, *Analytic functions* (Berlin: Springer) (1970)
- [22] Parthasarathy K R, *Introduction to Probability and Measure* (Delhi: Macmillan) (1977)
- [23] Reed M and Simon B, *Methods of modern mathematical physics, III, Scattering theory* (New York: Academic Press) (1979)
- [24] Simon B, Spectral analysis of rank one perturbations and applications, Lectures given at the Vancouver summer school of mathematical physics, August 10–14 (1993)
- [25] Sinha K B, On the theorem of M G Krein (Preprint) Univ. of Geneva, Geneva (1975)
- [26] Sveshnikov A and Tikhonov A, *Theory of functions of a complex variable* (Moscow: Mir Publishers) (1974)
- [27] Titchmarsh E C, *Introduction to the theory of Fourier Integrals* (2nd ed) (Oxford: University Press) (1975)
- [28] Voiculescu D, On a trace formula of M G Krein, (329–332), in *Operators in indefinite metric spaces, Scattering theory and other topics*, (eds. Helson, Nagy, Vascilescu, Voiculescu), (Basel: Birkhauser-Verlag) (1987)
- [29] Weidmann J, *Linear operators in Hilbert spaces* (New York: Springer Verlag) (1980)
- [30] Yafaev D R, *Mathematical scattering theory* (Providence, RI: American Mathematical Society) (1992)



# SUBJECT INDEX

- absolute summability
  - On the absolute matrix summability of Fourier series and some associated series 351
  - On absolute summability factors of infinite series 367
- algebraic group
  - Finite arithmetic subgroups of  $GL_n$ , III 201
  - Finite arithmetic subgroups of  $GL_n$ , III 201
- algebraic vector bundle
  - Vector bundles as direct images of line bundles 191
- almost periodicity
  - Inverse spectral theory for Jacobi matrices and their almost periodicity 777
- arithmetic-geometric mean
  - On the equation  $x(x + d_1) \dots (x + (k - 1)d_1) = y(y + d_2) \dots (y + (mk - 1)d_2)$  1
- auto regressive processes
  - Iterations of random and deterministic functions 263
- automorphic form
  - Zeta functions of prehomogeneous vector spaces with coefficients related to periods of automorphic forms 99
- automorphism
  - Some remarks on the Jacobian question 515
- ending shell model
  - Existence theory for linearly elastic shells 269
- Fermoulli numbers
  - Two remarkable doubly exponential series transformations of Ramanujan 245
- Fourier reduction
  - Positive values of non-homogeneous indefinite quadratic forms of type (1, 4) 557
- Fuchsian integrable functions
  - $L^1(\mu, X)$  as a complemented subspace of its bidual 421
- Fuchsian group
  - Non-surjectivity of the Clifford invariant map 49
- canonical class
  - Non-surjectivity of the Clifford invariant map 49
- Cesaro mean
  - On the absolute matrix summability of Fourier series and some associated series 351
- character sums
  - The number of ideals in a quadratic field 157
- Clifford invariant
  - Non-surjectivity of the Clifford invariant map 49
- Cohomology
  - Deformations of complex structures on  $\Gamma \backslash SL_2(C)$  389
- Combinatorial manifolds
  - Combinatorial manifolds with complementarity 385
- Commutator algebra
  - On  $N$ -body Schrödinger operators 667
- Complementarity
  - Combinatorial manifolds with complementarity 385
- Completely positive map
  - Kolmogorov's existence theorem for Markov processes in  $C^*$  algebras 253
- Counting function
  - The density of rational points on non-singular hypersurfaces 13
- Cusp forms
  - On Zagier's cusp form and the Ramanujan  $\tau$  function 93
- Decentralised learning algorithm
  - Absolutely expedient algorithms for learning Nash equilibria 279
- Dedekind's eta function
  - A note on a generalization of Macdonald's identities for  $A_l$  and  $B_l$  377
- Deformations
  - Deformations of complex structures on  $\Gamma \backslash SL_2(C)$  389
- Deligne's bounds
  - The density of rational points on non-singular hypersurfaces 13
- Differential subordination
  - Differential subordinations concerning starlike functions 397
- Diophantine equations
  - Row-reduction and invariants of Diophantine equations 549
- Direct image
  - Vector bundles as direct images of line bundles 191
- Distribution of zeros
  - On the zeros of a class of generalized Dirichlet series-XIV 167
- Doubly exponential series
  - Two remarkable doubly exponential series transformations of Ramanujan 245

|  |     |   |     |
|--|-----|---|-----|
| Elastic plate  |     | Generating functions  |     |
| Stresses in an elastic plate lying over a base due to strip-loading                                      | 425 | Iterations of random and deterministic functions  | 263 |
| Enss' method   |     | Global fields   |     |
| A conjecture for some partial differential operators on $L^2(R^n)$                                       | 705 | Reduction theory over global fields   | 207 |
| Equivalent forms   |     | Hierarchic control  |     |
| Positive values of non-homogeneous indefinite quadratic forms of type (1, 4)                             | 557 | Hierarchic control  | 295 |
| Exponential diophantine  |     | Hilbert-Schmidt integral operator   |     |
| On the equation $x(x + d_1) \dots (x + (k - 1)d_1) = y(y + d_2) \dots (y + (mk - 1)d_2)$                 | 1   | Extended Kac-Akhiezer formulae and the Fredholm determinant of finite section Hilbert-Schmidt kernels | 581 |
| Exponential sums   |     | Hilbert-Schmidt operators   |     |
| The number of ideals in a quadratic field  | 157 | The Hoffman-Wielandt inequality in infinite dimensions  | 483 |
| Finite arithmetic subgroup   |     | Hoffman-Wielandt inequality   |     |
| Finite arithmetic subgroups of $GL_n$ , III  | 201 | The Hoffman-Wielandt inequality in infinite dimensions  | 483 |
| Finite morphism  |     | Hyperbolic manifolds  |     |
| Vector bundles as direct images of line bundles  | 191 | The geometry and spectra of hyperbolic manifolds  | 715 |
| Fourier coefficients   |     | Hypersurfaces   |     |
| On Fourier coefficients of Maass cusp forms in 3-dimensional hyperbolic space                            | 77  | The density of rational points on non-singular hypersurfaces  | 13  |
| Fractional integral operator   |     | Ideals  |     |
| On composition of some general fractional integral operators   | 339 | The number of ideals in a quadratic field   | 157 |
| Fredholm determinant   |     | Infinite dimensions   |     |
| Extended Kac-Akhiezer formulae and the Fredholm determinant of finite section Hilbert-Schmidt kernels    | 581 | The Hoffman-Wielandt inequality in infinite dimensions  | 483 |
| Fubini-study metric  |     | Infinite series   |     |
| The Laplacian on algebraic threefolds with isolated singularities  | 435 | On absolute summability factors of infinite series  | 367 |
| Function fields  |     | Inverse theory  |     |
| Reduction theory over global fields  | 207 | Inverse spectral theory for Jacobi matrices and their almost periodicity                              | 777 |
| Functional calculus  |     | Iteration   |     |
| $L^p$ -Estimates for Schrödinger operators   | 653 | Iterations of random and deterministic functions  | 263 |
| Functional equation  |     | Jacobi forms  |     |
| Zeta functions of prehomogeneous vector spaces with coefficients related to periods of automorphic forms | 99  | Modular forms and differential operators  | 57  |
| GNS principle  |     | Jacobi matrices   |     |
| Kolmogorov's existence theorem for Markov processes in $C^*$ algebras                                    | 253 | Inverse spectral theory for Jacobi matrices and their almost periodicity                              | 777 |
| Gamma function   |     | Jacobian  |     |
| Two remarkable doubly exponential series transformations of Ramanujan                                    | 245 | Some remarks on the Jacobian question   | 515 |
| Gaussian quadrature  |     | Kac-Akhiezer formula  |     |
| Gaussian quadrature in Ramanujan's Second Notebook   | 237 | Extended Kac-Akhiezer formulae and the Fredholm determinant of finite section Hilbert-Schmidt kernels | 581 |
| Gaussian random walk   |     | Kloosterman sum   |     |
| On the structure of stable random walks  | 413 | On the Ramanujan-Petersson conjecture for modular forms of half-integral weight                       | 333 |
| General class of polynomials   |     | Koiter's model  |     |
| On composition of some general fractional integral operators   | 339 | Existence theory for linearly elastic shells  | 269 |

- Krein's theorem  
Spectral shift function and trace formula 819
- Kuznetsov theorem  
On Fourier coefficients of Maass cusp forms in 3-dimensional hyperbolic space 77
- $L^p$ -estimates  
 $L^p$ -Estimates for Schrödinger operators 653
- $L$ -ideals  
 $L^1(\mu, X)$  as a complemented subspace of its bidual 421
- Laplacian  
The Laplacian on algebraic threefolds with isolated singularities 435  
The geometry and spectra of hyperbolic manifolds 715
- Lattice  
Deformations of complex structures on  $\Gamma \backslash SL_2(C)$  389  
Rigidity problem for lattices in solvable Lie groups 495
- Lattices  
On a problem of G Fejes Tóth 137
- Lie groups  
Rigidity problem for lattices in solvable Lie groups 495
- Line bundle  
Vector bundles as direct images of line bundles 191
- Linearly elastic shells  
Existence theory for linearly elastic shells 269
- Linkage  
Bertini theorems for ideals linked to a given ideal 305
- Local Bertini theorem  
Bertini theorems for ideals linked to a given ideal 305
- Local zeta function  
Local zeta functions of general quadratic polynomials 177
- Maass cusp forms  
On Fourier coefficients of Maass cusp forms in 3-dimensional hyperbolic space 77
- Macdonald's multivariable identities  
A note on a generalization of Macdonald's identities for  $A_l$  and  $B_l$  377
- Markov process  
Kolmogorov's existence theorem for Markov processes in  $C^*$  algebras 253
- Membrane shell model  
Existence theory for linearly elastic shells 269
- Modular equations  
Modular equations and Ramanujan's Chapter 16, Entry 29 225
- Modular forms of half-integral weight  
On the Ramanujan-Petersson conjecture for modular forms of half-integral weight 333
- Multiple exponential sum  
The density of rational points on non-singular hypersurfaces 13
- Multiplicative properties  
Multiplicative arithmetic of finite quadratic forms over Dedekind rings 31
- Multivariable  $H$ -function  
On composition of some general fractional integral operators 339
- $N$ -body problems  
On  $N$ -body Schrödinger operators 667
- Nash blow up  
Bertini theorems for ideals linked to a given ideal 305
- Nash equilibria  
Absolutely expedient algorithms for learning Nash equilibria 279
- Neighbourhood of the critical line  
On the zeros of a class of generalised Dirichlet series-XIV 167
- Newton-Puiseux expansion  
Some remarks on the Jacobian question 515
- Non-compact knots  
On polynomial isotopy of knot-types 543
- Nonviscous  
A proof of Howard's conjecture in homogeneous parallel shear flows 593
- Norlund mean  
On the absolute matrix summability of Fourier series and some associated series 351
- Number fields  
Reduction theory over global fields 207
- Optimality system  
Hierarchic control 295
- Partial differential operators  
A conjecture for some partial differential operators on  $L^2(R^n)$  705
- Plane curves  
Some remarks on the Jacobian question 515
- Poisson summation  
Two remarkable doubly exponential series transformations of Ramanujan 245
- Polynomial isotopy  
On polynomial isotopy of knot-types 543
- Prehomogeneous vector space  
Zeta functions of prehomogeneous vector spaces with co-efficients related to periods of automorphic forms 99
- Projective variety  
Vector bundles as direct images of line bundles 191

|   |     |   |     |
|---|-----|---|-----|
| Quadratic field   |     | Scattering theory   |     |
| The number of ideals in a quadratic field                                       | 157 | Scattering theory for Stark hamiltonians                                    | 599 |
| Quadratic forms   |     | A conjecture for some partial differential operators on $L^2(\mathbb{R}^n)$ | 705 |
| Multiplicative arithmetic of finite quadratic forms over Dedekind rings         | 31  | Schatten $p$ -norms   |     |
| Non-surjectivity of the Clifford invariant map                                  | 49  | The Hoffman-Wielandt inequality in infinite dimensions                      | 483 |
| Finite arithmetic subgroups of $GL_n$ , III                                     | 201 | Schrödinger operators   |     |
| Positive values of non-homogeneous indefinite quadratic forms of type (1,4)     | 557 | Scattering theory for Stark hamiltonians                                    | 599 |
| Quadratic polynomials   |     | $L^p$ -Estimates for Schrödinger operators                                  | 653 |
| Local zeta functions of general quadratic polynomials                           | 177 | Self-adjointness  |     |
| $r$ -fold basic elements  |     | The Laplacian on algebraic threefolds with isolated singularities           | 435 |
| Bertini theorems for ideals linked to a given ideal                             | 305 | Sequences   |     |
| Ramanujan   |     | Rearrangements of bounded variation sequences                               | 373 |
| Gaussian quadrature in Ramanujan's Second Notebook                              | 237 | Series approximations   |     |
| Ramanujan tau function  |     | Gaussian quadrature in Ramanujan's Second Notebook                          | 237 |
| On Zagier's cusp form and the Ramanujan $\tau$ function                         | 93  | Shear flows   |     |
| Ramanujan-Petersson conjecture  |     | A proof of Howard's conjecture in homogeneous parallel shear flows          | 593 |
| On the Ramanujan-Petersson conjecture for modular forms of half-integral weight | 333 | Shear stresses  |     |
| Random maps   |     | Stresses in an elastic plate lying over a base due to strip-loading         | 425 |
| Iterations of random and deterministic functions                                | 263 | Shear surface loads   |     |
| Rational points   |     | Stresses in an elastic plate lying over a base due to strip-loading         | 425 |
| The density of rational points on non-singular hypersurfaces                    | 13  | Siegel theorem  |     |
| Rearrangements  |     | Multiplicative arithmetic of finite quadratic forms over Dedekind rings     | 31  |
| Rearrangements of bounded variation sequences                                   | 373 | Singular locus  |     |
| Reduction theory  |     | The density of rational points on non-singular hypersurfaces                | 13  |
| Reduction theory over global fields   | 207 | Singularities   |     |
| Resolvent estimates   |     | The Laplacian on algebraic threefolds with isolated singularities           | 435 |
| On $N$ -body Schrödinger operators  | 667 | Smooth base   |     |
| Rigid base  |     | Stresses in an elastic plate lying over a base due to strip-loading         | 425 |
| Stresses in an elastic plate lying over a base due to strip-loading             | 425 | Spectral shift function   |     |
| Rigidity problem  |     | Spectral shift function and trace formula                                   | 819 |
| Rigidity problem for lattices in solvable Lie groups                            | 495 | Spectral theory   |     |
| Rings of automorphs   |     | Scattering theory for Stark hamiltonians                                    | 599 |
| Multiplicative arithmetic of finite quadratic forms over Dedekind rings         | 31  | Sphere (balls)  |     |
| Rogers-Ramanujan functions  |     | On a problem of G Fejes Toth  | 137 |
| Modular equations and Ramanujan's Chapter 16, Entry 29                          | 225 | Stable processes  |     |
| Row-reduction   |     | Iterations of random and deterministic functions                            | 263 |
| Row-reduction and invariants of Diophantine equations                           | 549 | Stable random walks   |     |
| $S$ -matrices   |     | On the structure of stable random walks                                     | 413 |
| On $N$ -body Schrödinger operators  | 667 | Stackleberg terminology   |     |
|   |     | Hierarchic control  | 295 |
|   |     | Starlike and convex functions   |     |
|   |     | Differential subordinations concerning starlike functions                   | 397 |
|   |     | Summability factors   |     |
|   |     | On absolute summability factors of infinite series                          | 367 |



|   |     |  |     |
|---|-----|--|-----|
| symplectic structures   |     | Univalent  |     |
| Symplectic structures on locally compact abelian groups and polarizations       | 217 | Differential subordinations concerning starlike functions  | 397 |
| thinnest arrangements   |     | Vector measures  |     |
| On a problem of G Fejes Toth  | 137 | $L^1(\mu, X)$ as a complemented subspace of its bidual   | 421 |
| three folds   |     | Vertex operator algebras   |     |
| The Laplacian on algebraic threefolds with isolated singularities               | 435 | Modular forms and differential operators   | 57  |
| three-dimensional hyperbolic space  |     | Zariski density  |     |
| On Fourier coefficients of Maass cusp forms in 3-dimensional hyperbolic space   | 77  | Rigidity problem for lattices in solvable Lie groups   | 495 |
| trace formula   |     | Zeta function  |     |
| Spectral shift function and trace formula                                       | 819 | Zeta functions of prehomogeneous vector spaces with coefficients related to periods of automorphic forms | 99  |
| triangular matrix   |     |  |     |
| On the absolute matrix summability of Fourier series and some associated series | 351 |  |     |

|  |     |   |     |
|--|-----|---|-----|
| Aaronson Jon   |     | Garg Nat Ram  |     |
| On the structure of stable random walks                                      | 413 | see Sharma Raj Kumar  | 425 |
| Abhyankar Shreeram S   |     | Gupta K C   |     |
| Some remarks on the Jacobian question  | 515 | On composition of some general fractional integral operators                    | 339 |
| Andrews George E   |     | Hafner James Lee  |     |
| Modular equations and Ramanujan's Chapter 16, Entry 29                       | 225 | see Berndt Bruce C  | 245 |
| Andrianov Anatoli  |     | Hashim Ashwaq   |     |
| Multiplicative arithmetic of finite quadratic forms over Dedekind rings      | 30  | On Zagier's cusp form and the Ramanujan $\tau$ function                         | 93  |
| Antony Anand J   |     | Heath-Brown D R   |     |
| Inverse spectral theory for Jacobi matrices and their almost periodicity     | 777 | The density of rational points on non-singular hypersurfaces                    | 13  |
| Askey Richard  |     | Hirschowitz A   |     |
| Gaussian quadrature in Ramanujan's Second Notebook                           | 237 | Vector bundles as direct images of line bundles                                 | 191 |
| Athreya K B  |     | Hislop Peter D  |     |
| Iterations of random and deterministic functions                             | 263 | The geometry and spectra of hyperbolic manifolds                                | 715 |
| Balasubramanian R  |     | Huxley M N  |     |
| On the zeros of a generalised Dirichlet series-XIV                           | 167 | The number of ideals in a quadratic field                                       | 157 |
| Bambah R P   |     | Igusa Jun-ichi  |     |
| On a problem of G Fejes Toth   | 137 | Local zeta functions of general quadratic polynomials                           | 177 |
| Banerjee Mihir B.  |     | Isozaki Hiroshi   |     |
| A proof of Howard's conjecture in homogeneous parallel shear flows           | 593 | On $N$ -body Schrödinger operators  | 667 |
| Berndt Bruce C   |     | Jensen Arne   |     |
| Two remarkable doubly exponential series transformations of Ramanujan        | 245 | Scattering theory for Stark hamiltonians  | 599 |
| Bhatia Rajendra  |     | Kanwar Vinay  |     |
| The Hoffman-Wielandt inequality in infinite dimensions                       | 483 | see Banerjee Mihir B  | 593 |
| Bör Huseyin  |     | Kitaoka Yoshiyuki   |     |
| On absolute summability factors of infinite series                           | 367 | Finite arithmetic subgroups of $GL_n$ , III                                     | 201 |
| Ciarlet Philippe G   |     | Köhnen Winfried   |     |
| Existence theory for linearly elastic shells                                 | 269 | On the Ramanujan-Petersson conjecture for modular forms of half-integral weight | 333 |
| Datta Basudeb  |     | Krishna M   |     |
| Combinatorial manifolds with complementarity                                 | 385 | see Sinha Kalyan B (Foreword)   | 597 |
| Dumir V C  |     | see Antony Anand J  | 777 |
| Positive values and non-homogeneous indefinite quadratic forms of type (1,4) | 557 | Lions J L   |     |
| Elsner Ludwig  |     | Hierarchic control  | 295 |
| see Bhatia Rajendra  | 483 | Mohapatra A N   |     |
|  |     | see Sinha Kalyan B  | 819 |
|  |     | Murty M Ram   |     |
|  |     | see Hashim Ashwaq   | 93  |

- Muthuramalingam Pl.  
A conjecture for some partial differential operators  
on  $L^2(R^n)$  705
- Nakamura Shu  
 $L^p$ -Estimates for Schrödinger operators 653
- Narasimhan M S  
see Hirschowitz A 191
- Parimala R  
Non-surjectivity of the Clifford invariant map 49
- Parthasarathy K R  
see Rajarama Bhat B V 253
- Pati Vishwambhar  
The Laplacian on algebraic threefolds with  
isolated singularities 435
- Phansalkar V V  
Absolutely expedient algorithms for learning  
Nash equilibria 279
- Ponnusamy S  
Differential subordinations concerning starlike  
functions 397
- Raghavan S  
On Fourier coefficients of Maass cusp forms in  
3-dimensional hyperbolic space 77
- Rajan C S  
Deformations of complex structures on  $\Gamma \backslash SL_2(C)$   
389
- Rajarama Bhat B V  
Kolmogorov's existence theorem for Markov  
processes in  $C^*$  algebras 253
- Ramachandra K  
see Balasubramanian R 167
- Raman S Ganapathi  
Extended Kac-Akhiezer formulae and the Fred-  
holm determinant of finite section Hilbert-  
Schmidt kernels 581
- Rao R Ranga  
Symplectic structures on locally compact abelian  
groups and polarizations 217
- Rao R Vittal  
see Raman S Ganapathi 581
- Rao T S S R K  
 $L^1(\mu, X)$  as a complemented subspace of its bidual  
421
- Ray B K  
On the absolute matrix summability of Fourier  
series and some associated series 351
- Sahoo A K  
see Ray B K 351
- Saradha N  
On the equation  $x(x+d_1)\dots(x+(k-1)d_1)=$   
Sarigol Mehmet Ali  
Rearrangements of bounded variation sequences 373
- Sastry P S  
see Phansalkar V V 279
- Sato Fumihito  
Zeta functions of prehomogeneous vector spaces  
with coefficients related to periods of automor-  
phic forms 99
- Sehmi Ranjeet  
see Dumir V C 557
- Sengupta J  
see Raghavan S 77
- Shandil R G  
see Banerjee Mihir B 593
- Sharma Raj Kumar  
Stresses in an elastic plate lying over a base due  
to strip-loading 425
- Shorey T N  
see Saradha N 1
- Shukla Rama  
On polynomial isotopy of knot-types 543
- Sinha Kalyan B  
Foreword 597  
Spectral shift function and trace formula 819
- Soni R C  
see Gupta K C 339
- Springer T A  
Reduction theory over global fields 207
- Sridharan R  
see Parimala R 49
- Starkov A N  
Rigidity problem for lattices in solvable Lie  
groups 495
- Sthanumoorthy N  
A note on a generalization of Macdonald's  
identities for  $A_l$  and  $B_l$  377
- Tamba M  
see Sthanumoorthy N 377
- Thathachar M A L  
see Phansalkar V V 279
- Vijayalaxmi Trivedi  
Bertini theorems for ideals linked to a given ideal  
305
- Watt N  
see Huxley M N 157
- Wildberger N J  
Row-reduction and invariants of Diophantine  
equations 549
- Woods A C  
see Bambah R P 137
- Zagier Don